THE MOUSE GLANDULAR KALLIKREIN GENE FAMILY

BY

ANTHONY JOHN MASON

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY AT THE
AUSTRALIAN NATIONAL UNIVERSITY,
CANBERRA

JULY, 1983

STATEMENT

All the work described in the thesis was performed by myself except where due reference is made in the text. No material in this thesis has been presented for any other degree or diploma.

Anthony J. Mason

ACKNOWLEDGEMENTS

I would like to thank my two supervisors Rob Richards and John Shine for their constant encouragement, advice and friendly support during the past three years. Somehow they managed to convince me that it was worth sequencing 9500 bp of DNA. John Shine has also built up my moral fibre by insisting on taking five dollars off me every time Hawthorn beat Collingwood.

I am also extremely grateful to the many members and visitors of the Shine laboratory for their friendship and encouragement. I would especially like to thank Bronwyn Evans, Nancy Cooke, Nick Deacon, George Miklos, Kieran Scott, John Watson and Peter Whitfeld for their readiness to discuss problems relating to this project. I wish to thank Professor John Pateman for permission to work in the Department of Genetics, and to Parry Monckton for encouraging me to come to Canberra for my Ph.D. I have been financially supported by a Commonwealth Postgraduate Research Award during the course of my studies.

I appreciate the efforts of John Watson, Barbara Van Leeuwen and Bronwyn Evans in proof-reading this thesis, Rob Evans, Judy Hardman and Yvonne Hort for help with photography, Angela Hanbury-Sparrow and Val Rawlings for typing this thesis.

This thesis is a tribute to the faith my parents, family, and friend Leo have shown in me over many years.

ABSTRACT

The work in this thesis describes the cloning and initial characterisation of a large gene family whose members encode specific proteases involved in the processing of biologically active peptides. This family has been termed the glandular kallikrein multi-gene family since all its members encode serine proteases closely related to glandular kallikrein. Glandular kallikrein enzyme activity has been implicated by *in vivo* and *in vitro* studies in a wide variety of essential physiological functions, in addition to its role as a kininogenase. The use of molecular techniques has shown that the glandular kallikrein gene family comprises 25-30 highly homologous but distinct genes, all of which are closely linked on chromosome seven in the mouse.

The 9.5 kilobase (kb) nucleotide sequence of a mouse genomic clone (λMSP-1) which contains one complete kallikrein gene (mGK-1) and the 3' end of another (mGK-2) has been determined. The two genes on λMSP-1 are encoded on the same DNA strand; the first gene (mGK-1) is complete and consists of five exons and four introns, together spanning 4.5 kb. The 3' end of another gene (mGK-2) is located 3.7 kb upstream from the 5' end of mGK-1. The mGK-1 gene has been shown to be transcriptionally active in the male mouse submaxillary gland.

The isolation and partial characterisation of other genomic kallikrein clones has demonstrated that the kallikrein genes are frequently tightly linked and are highly conserved in their DNA sequence and structural organisation. Differences in the coding potential of these genes and the amino acid sequences of other known kallikreins seem to be functionally related to the substrate specificity of the different enzymes.

In order to study the expression of the mGK-1 gene, a synthetic 19-mer homologous to a portion of mGK-1 was designed. This oligonucleotide was found to hybridise only to mGK-1 sequences. This method offers a means of investigating the tissue-specific expression of individual members of the kallikrein gene family.

ABBREVIATIONS

AMV:	avian myeloblastosis virus
Ap^r:	ampicillin resistant
Bis:	N,N'methylene bisacrylamide
cDNA:	complementary DNA
Ci:	Curie
ddNTP:	dideoxy-nucleotide-5'-triphosphate
dNTP:	2'-deoxy-nucleotide-5'-triphosphate
DNAse:	deoxyribonuclease
ds:	double-stranded
EDTA:	ethylenediaminetetracetic acid, disodium salt
Hepes:	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid
IPTG:	isopropyl-ß-D-thiogalactopyranoside
kb:	kilobase
NaOAc:	sodium acetate
pfu:	plaque forming unit
RNAse:	ribonuclease
ss:	single-stranded
SSC:	standard saline citrate
Tc^r:	tetracycline resistant
Tris:	tris(hydroxymethyl)-aminomethane
UV:	ultraviolet
X-gal:	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside

PAPERS PUBLISHED

Some of the work described in this thesis has been published in the following papers:

1. Mason, A.J., B.A. Evans, D.R. Cox, J. Shine and R.I. Richards (1983). Structure of the mouse kallikrein gene family suggests a role in specific processing of biologically active peptides. Nature 303, 300-307.

2. Richards, R.I., D.F. Catanzarro, A.J. Mason, B.J. Morris, J.D. Baxter and J. Shine (1982). Mouse glandular kallikrein genes. J. Biol. Chem. 257, 2758-2761.

TABLE OF CONTENTS

CHAPTER ONE

GENERAL INTRODUCTION

Eukaryotic protein biosynthesis is an extremely complex process involving a multitude of steps, most of which can be envisaged as points at which the expression of a particular phenotype may be regulated (Brown, 1981). Some of the regulatory mechanisms which operate on gene expression at the transcriptional, post-transcriptional and translational stages of protein biosynthesis have been identified and are the subject of recent reviews (Darnell, 1982; Brown, 1981). Furthermore one can speculate that several types of post-translational control may operate, which also affect the final concentration of a particular protein in a cell. Following translation, proteins undergo many significant modifications before they appear in their biologically active/functional form. These include the alteration of specific amino acids by phosphorylation, glycosylation and acylation mechanisms (Uy and Wold, 1977). In addition many biologically active peptides are initially synthesised as precursors which require a highly specific peptide cleavage event to occur before their activity can be fully realised (Neurath and Walsh, 1976). These final maturation steps represent another control point at which the expression of a eukaryotic gene may be influenced.

The work described in this thesis involves the characterisation of a multi-gene family, members of which are responsible for the post-translational maturation of certain growth factors. The family has been termed the

kallikrein gene family since all of its members are closely related to glandular kallikrein. The following literature survey provides a background to the work described in this thesis.

1.1 Post-translational limited proteolysis

Since the initial discovery that proteolytic digestive enzymes such as trypsin and chymotrypsin are activated by the selective cleavage of a peptide bond (Neurath and Dixon, 1957), it has become apparent that similar mechanisms operate in many fundamental biological processes. Limited proteolysis plays an important role in blood coagulation (Davie and Fujikawa, 1975), blood pressure regulation (Orstavik, 1980), the complement pathway (Hugli, 1975) and polypeptide hormone production (Docherty and Steiner, 1982).

The remainder of this chapter summarises what is currently known about the enzymes involved in polypeptide hormone maturation. Two control mechanisms have evolved to regulate the action of enzymes responsible for limited proteolysis as exemplified by the activation of trypsinogen. Trypsinogen, which is synthesised in the acinar cells of the pancreas, is initially activated by the highly specific protease, enterokinase, which is secreted from the brush border of the small intestine (Maroux *et al.*, 1971). Thus, by compartmentalising the substrate and the enzyme to different cells, the activity of trypsin

is restricted to the digestive tract where it performs its primary degradative functions. This example also demonstrates the specificity of such limited proteolysis, since only one peptide bond is cleaved in trypsinogen and no other protein has yet been reported to be a substrate for enterokinase (Maroux et al., 1971). These control mechanisms of compartmentalisation and specificity may also be used to control the enzymes involved in the maturation of polypeptide hormones as detailed in the following sections.

1.2 Polypeptide hormone precursors

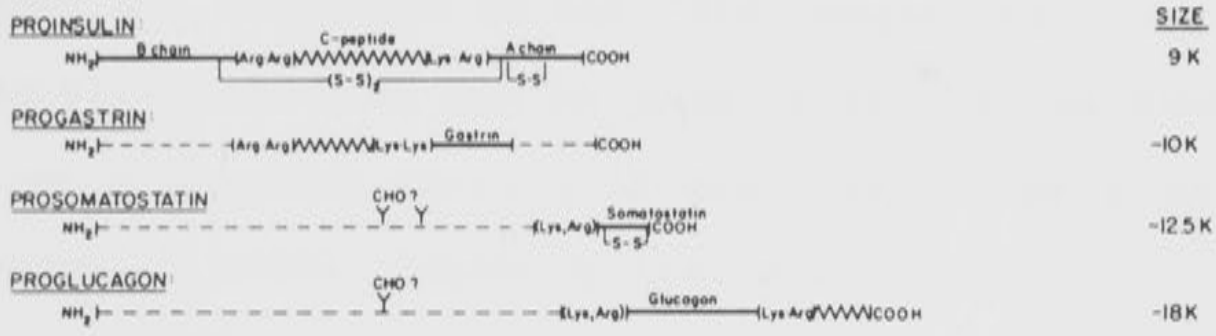
Peptide hormones fall into two categories according their biosynthetic pathways. The prolactin (Cooke et al., 1981), placental lactogen (Shine et al., 1977), growth hormone (Seeburg et al., 1977) family and the glycoprotein hormone family (Fiddes and Goodman, 1979, 1980) are examples of hormones initially synthesised with a very hydrophobic NH_2-terminal prepeptide. This signal sequence, common to most secretory proteins (Chrétien and Seidah, 1981) assists in the initial microsomal segregation of peptides destined for secretion. This is achieved by the signal recognition particle, which mediates the specific attachment of polysomes involved in the synthesis of secretory proteins to the microsomal membrane of the endoplasmic reticulum (Walter and Blobel, 1981). During the process of protein translocation across the membrane,

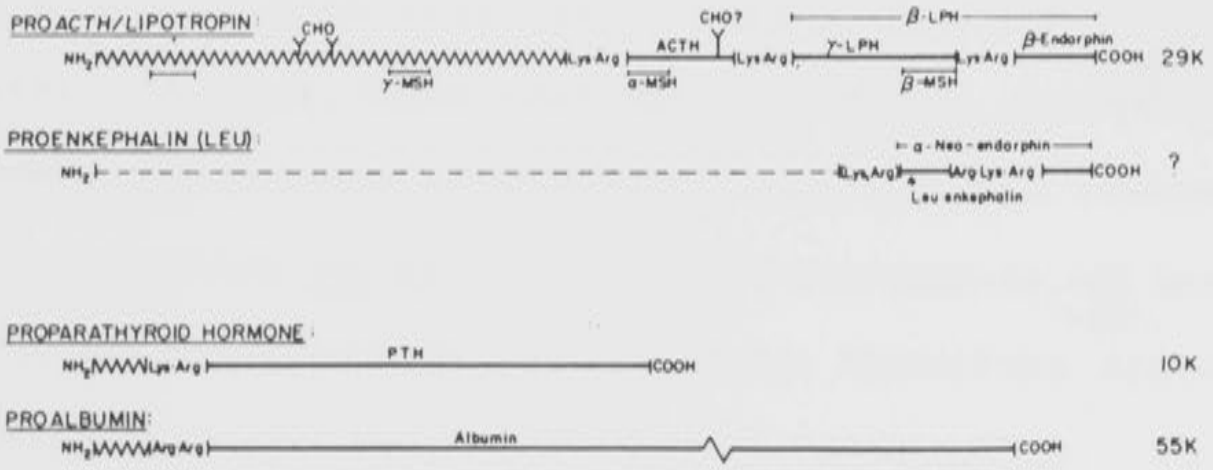
the prepeptide sequence is removed (Walter and Blobel, 1980). The above-mentioned hormones, after membrane transport and packaging are present in their physiologically active form ready for export.

However, the vast majority of hormones fall into a different category, since in addition to a prepeptide which is lost during membrane transport, they possess extra 'pro'-peptide sequences which are not required for biological activity. Such hormones include adrenocorticotropin (ACTH) - β-lipotropin (β-LPH) (Nakanashi et al., 1979), insulin (Steiner and Oyer, 1967; Ullrich et al., 1977), relaxin (Hudson et al., 1981a), somatostatin (Hobart et al., 1980), vasopressin (Land et al., 1982), oxytocin (Land et al., 1983) and the enkephalin precursors (Kakidani et al., 1982; Noda et al., 1982).

As a result of molecular cloning studies, it has recently become clear that many of the smaller peptide hormones are initially synthesised as part of a common precursor containing several different hormone sequences, often peptides which are regulated by the same physiological stimuli, such as the ACTH-β-LPH precursor (Nakanashi et al., 1979). An examination of the primary structure of hormone precursors (Figure 1.1) reveals that cleavage at one or a number of peptide bonds is required to release the biologically active component of the precursor. It would appear therefore that limited

Figure 1.1 Diagrammatic representation of the structure of polypeptide precursors (diagram reproduced from Steiner, 1982). The final biosynthetic products are represented as dark lines, and additional 'pro' sequences as wavy or broken lines. More detailed information on the structure of the ACTH and enkephalin precursor forms has been published elsewhere (Nakanashi *et al*. 1979; Kakandani *et al*., 1982; Noda *et al*., 1982).


PROINSULIN
SIZE
NH2
B chain
C-peptide
A chain
COOH
(S-S)2
S-S
9 K
PROGASTRIN
NH2
(Arg Arg)
(Lys Lys)
Gastrin
COOH
~10K
PROSOMATOSTATIN
CHO ?
(Lys, Arg)
Somatostatin
COOH
S-S
~12.5 K
PROGLUCAGON
CHO ?
(Lys, Arg)
Glucagon
(Lys Arg)
COOH
~18K


PROACTH/LIPOTROPIN
CHO
CHO?
β-LPH
ACTH
γ-LPH
β-Endorphin
NH2
(Lys Arg)
COOH
29K
γ-MSH
α-MSH
β-MSH
PROENKEPHALIN (LEU)
α-Neo-endorphin
NH2
(Lys Arg)
Arg Lys Arg
COOH
?
Leu enkephalin
PROPARATHYROID HORMONE
NH2
(Lys Arg)
PTH
COOH
10K
PROALBUMIN
NH2
(Arg Arg)
Albumin
COOH
55K

proteolysis plays an important role in hormone biosynthesis, and thus may be an important step in regulating hormone activity. For this reason an examination of the enzymes involved is essential for a complete understanding of hormone biosynthesis.

1.3 Enzymes implicated in polypeptide hormone maturation

A comparison of the proforms of various hormones reveals little in common with regard to the position, length and structure of the 'pro'-sequences. One generalisation that can be made is that sites cleaved during processing consist of basic amino acids usually present in pairs, commonly lysine-arginine (see Figure 1.1). Studies on an abnormal proinsulin, in which one of the cleavage site arginine residues has been changed to a neutral residue, show that this mutation prevents the correct processing of the insulin precursor (Gabbay et al., 1979; Robbins et al., 1981). The presence of basic amino acids at the cleavage points would therefore appear to be a necessary requirement for correct maturation. It has been suggested that this requirement for basic amino acids points to the involvement of a tryptic-like enzyme(s) in hormone processing (Steiner, 1982). In addition a carboxypeptidase-like activity would be required to remove the exposed basic residue after the initial cleavage has occurred (Steiner, 1982).

Several studies have implicated the involvement of

both acid thiol and serine proteases in prohormone processing.

Since cleavage probably occurs at the stage of secretory processing near the Golgi apparatus and continues in the secretory granules (Sun et al., 1973; Steiner et al., 1974), extracts from these granules have been used to isolate potential processing enzymes. Extracts from the secretory granules of anglerfish and rat pancreatic islets have been shown to contain an activity which converts proinsulin, prosomatostatin and proglucagon into their respective active configurations (Fletcher et al., 1980; Fletcher et al., 1981; Docherty et al., 1982). By the use of specific inhibitors, these studies were able to demonstrate that the 'converting activity' shared many properties with acid thiol proteases. This activity was similar but clearly distinct from the well characterised acid thiol protease cathepsin B.

An intriguing situation is the differential processing of the multi-hormone precursor pro-opiomelanocortin (POMC) in the anterior and intermediate lobes of the pituitary. The predominant products synthesised in the anterior lobe are ACTH and β-LPH while the intermediate lobe produces mainly α-melanotropin (α-MSH) and β-endorphin (Chrétien and Siedah, 1981). One possible explanation for this differential expression of the component peptides of POMC is that different processing enzymes exist in each of the pituitary lobes. Loh and Gainer (1982) found an enzyme

activity in the secretory granules of rat intermediate lobe which was capable of converting toad POMC into ACTH, α-MSH and β-LPH but incapable of further processing β-LPH to give β-endorphin. Inhibitor studies again indicated that this activity was consistent with that of an acid thiol protease distinct from cathepsin B or D. Conversely other workers have demonstrated that a highly specific kallikrein-like serine protease called tonin can also cleave ACTH from the POMC precursor. This enzyme, present in large amounts in the rat submaxillary gland has also been found in the anterior lobe of the rat pituitary (Chrétien and Seidah, 1981). Clearly more detailed biochemical and physiological studies are needed before any of the above enzyme activities can be proven to be responsible for hormone maturation.

Although very little is known about the enzyme(s) responsible for the specific processing of many hormonal peptides, a considerable amount of data has accumulated concerning the processing of nerve growth factor (NGF) and epidermal growth factor (EGF). Unlike the instances discussed above, the enzymes which process NGF and EGF remain bound to their specific substrates after the cleavage event facilitating the isolation and study of these particular processing enzymes (Frey et al., 1979; Berger and Shooter, 1977). Detailed biochemical studies have revealed that these enzymes are highly homologous serine proteases closely related to glandular kallikrein

(Bothwell et al., 1979). The remainder of this chapter will be devoted to a description of the properties of these enzymes and the evidence which suggests that they are encoded by members of a multi-gene family.

1.4 Structure of the EGF and NGF complexes

The male mouse submaxillary gland, being the richest source known for EGF and NGF, has been the tissue of choice for detailed studies of the biosynthesis of these growth factors (Frey et al., 1979; Berger and Shooter, 1977). EGF and NGF have been isolated from this gland as high molecular weight (HMW) complexes. The EGF complex of 79,000 d, known as HMW-EGF, is composed of two EGF peptide chains, two molecules of binding protein (EGF-BP) and two zinc ions (Taylor et al., 1970). The growth-promoting activities of EGF, which include the accelerated proliferation and differentiation of skin tissue (Birnbaum et al., 1976), corneal epithelial tissue (Ho et al., 1974) and lung epithelium (Catterton et al., 1979), are attributed solely to the EGF peptide. The EGF-BP is a trypsin-like arginyl esteropeptidase (Taylor et al., 1970).

High molecular weight NGF; (7S NGF; 130,000 d) is a complex of three subunits (α, β, γ) having the stoichiometry $\alpha_2\gamma_2\beta$ (Server and Shooter, 1976). The stability of this complex is dependent on the presence of zinc ions (Pattison and Dunn, 1975). The β-NGF peptide which is entirely responsible for the biological activity

of NGF (Greene et al., 1971), is composed of two identical peptide chains each containing 118 amino acids (Angeletti and Bradshaw, 1971). NGF activity has been shown to be essential for the survival and differentiation of sympathetic and sensory neurons (Levi-Montalcini and Angeletti, 1968; Levi-Montalcini, 1976) as well as the survival of adrenal chromaffin cell precursors (Aloe and Levi-Montalcini, 1979). The γ-subunit is a trypsin-like arginyl esteropeptidase with very similar properties to EGF-BP (Greene et al., 1969; Server and Shooter, 1976). Although strikingly similar, γ-NGF and EGF-BP will not substitute for each other in complex formation (Server and Shooter, 1976). No activity has yet been assigned to the α-subunit of NGF (Stach et al., 1980).

The finding that the growth-promoting activity and the arginyl esteropeptidase activity both exist within the same intracellular granules, discounts any suggestion that these complexes are artifacts of purification (Pasquini et al., 1974). Their biological relevance is also emphasised by their presence predominantly as HMW-complexes in mouse saliva (Burton et al., 1978). Studies have shown that the 7S NGF complex must dissociate before any biological activity is achieved (Stach and Shooter, 1980; Harris-Warrick et al., 1980). The biological effects of NGF and EGF have been recently reviewed (Carpenter and Cohen, 1979; Thoenen and Barde, 1980). The following sections will be limited to a discussion of their biosynthesis.

1.5 Biosynthesis of EGF (Frey et al., 1979)

The biologically active form of EGF, isolated from the HMW-EGF complex, is a peptide of 53 amino acids having a molecular weight M_r of 6045 d (Savage et al., 1972). A precursor to EGF, of molecular weight 9000 d was demonstrated by pulse-chase experiments performed on isolated mouse submaxillary glands. This M_r9000 species was isolated in its native form from submaxillary gland extracts, and shown by chymotryptic peptide mapping to contain all the peptides characteristic of EGF except for a modified COOH-terminal peptide. This indicates that the M_r3000 propeptide is located at the COOH-terminal end of proEGF.

Incubation of the M_r9000 species with purified EGF-BP, at a 1:1 molar ratio, resulted in its complete conversion to the M_r6045 EGF species and the formation of the HMW-EGF complex. If EGF-BP was added to the M_r9000 species at a molar ratio of 0.5:1, only 50% of the M_r9000 species was converted to EGF. The reaction is thus stoichiometric rather than catalytic. The absolute specificity of this reaction was demonstrated by the inability of a five hundred fold excess of γ-NGF, β-NGF endopeptidase or trypsin to cleave the M_r9000 species. Thus EGF-BP is capable of processing proEGF and participating in the subsequent complex formation (Figure 1.2).

The pulse chase experiments also indicated that there is an even larger precursor form of EGF which is probably

not solubilised under the conditions used or is not precipitated with EGF antibodies. The cloning of the mRNA for mouse EGF has indicated that the initial precursor form of EGF is considerably larger than the M_r9000 species (Ullrich, A., pers. comm.). It remains to be seen whether EGF-BP is responsible for the processing of this earlier precursor, but it is clearly responsible for the processing of the M_r9000 species to EGF.

1.6 Biosynthesis of NGF (Berger and Shooter, 1977)

The biosynthetic events believed to be involved in the formation of the 7S NGF complex are outlined in Figure 1.3. Pulse chase experiments performed on isolated submaxillary glands demonstrated an immediate M_r 22,000 precursor of β-NGF. The γ-subunit of NGF is capable of cleaving this peptide to release individual β-NGF chains of 13,260 d. Unexpectedly EGF-BP was also capable of cleaving proβ-NGF correctly. However this seems unlikely to occur *in vivo* since γ-NGF would still be needed for the formation of the 7S NGF complex; an event in which EGF-BP is incapable of participating (Server and Shooter, 1976). The probable explanation for this result as advanced by Frey *et al.*, (1979), is that the proβ-NGF used was not in its native configuration since it was isolated from a denaturing gel. In contrast, experiments on proEGF (see 1.4) were performed on a protein isolated directly from a

Figure 1.2. Diagrammatic representation of the postulated biosynthetic events in the formation of the high molecular weight EGF complex (reproduced from Frey et al., 1979). More recent experimental data indicate that proEGF is a considerably larger peptide than that represented in this diagram (Ullrich, A., pers. comm.).

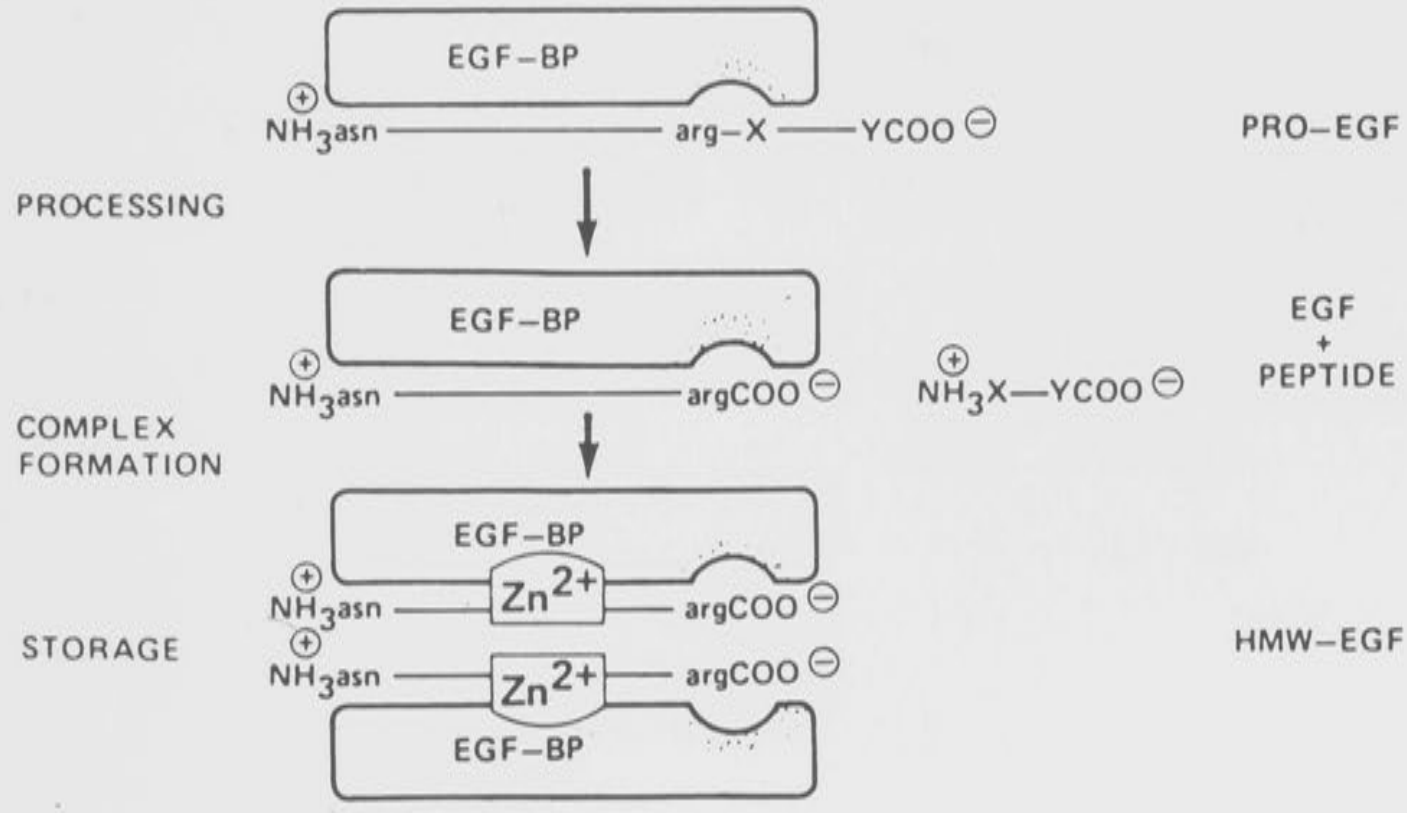
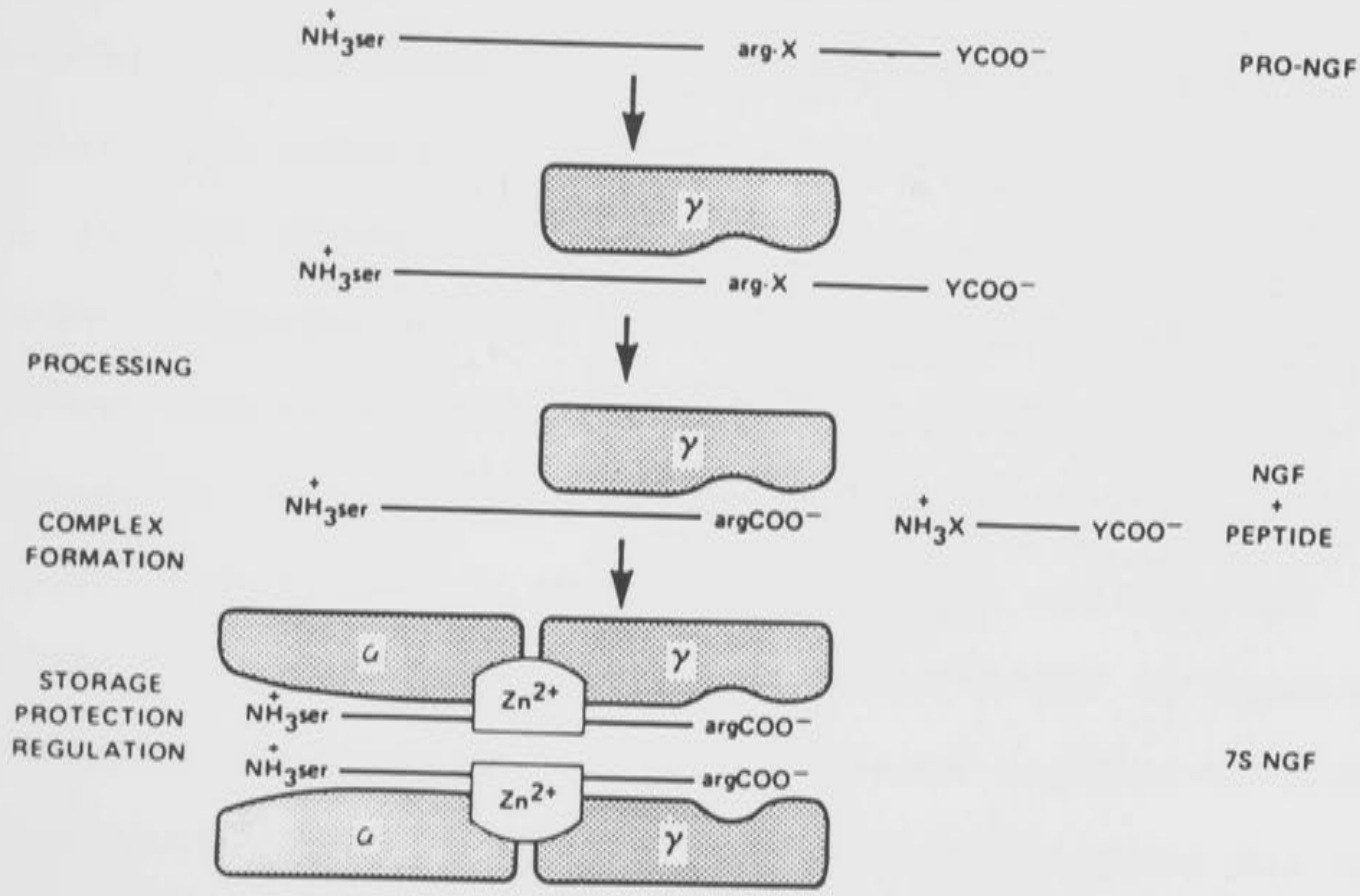

EGF–BP
NH3asn
arg–X
YCOO
PRO–EGF
PROCESSING
EGF–BP
NH3asn
argCOO
NH3X—YCOO
EGF
+
PEPTIDE
COMPLEX
FORMATION
EGF–BP
NH3asn
Zn2+
argCOO
STORAGE
NH3asn
Zn2+
argCOO
EGF–BP
HMW–EGF

Figure 1.3 Biosynthesis of NGF and its high molecular weight complex, 7S NGF (Reproduced from Shooter et al., 1982). As detailed in 1.6, the precursor form of NGF contains additional NH_2-terminal 'pro' sequences than illustrated here (Scott et al., 1983).


NH_3^+ser — arg·X — $YCOO^-$
PRO-NGF
γ
NH_3^+ser — arg·X — $YCOO^-$
PROCESSING
γ
COMPLEX FORMATION
NH_3^+ser — $argCOO^-$
NH_3^+X — $YCOO^-$
NGF + PEPTIDE
G
γ
STORAGE PROTECTION REGULATION
Zn^{2+}
NH_3^+ser — $argCOO^-$
NH_3^+ser — $argCOO^-$
Zn^{2+}
7S NGF
G
γ

gland homogenate (Frey et al., 1979). The α-subunit of NGF may also confer some specificity on the reaction.

The recent molecular cloning of the mRNA for β-NGF (Scott et al., 1983) has demonstrated that the initial precursor for β-NGF is larger than the M_r 22,000 precursor protein described by Berger and Shooter (1976). The nucleotide sequence of this clone predicts a 307 amino acid polypeptide which contains NGF (amino acids 188-305). There are three potential basic cleavage sites within the NH_2-terminal extension of NGF, at positions 115-116, 144-147, 186-187 respectively; and an Arg-Arg pair at position 305-306, adjacent to the carboxy-terminus. Whether cleavage occurs at all of these sites is unknown, but cleavage at residues 115-116 would result in a precursor product of M_r22,000 as observed by Berger and Shooter (1976). The authors suggest that since γ-NGF is capable of binding to the carboxy-terminal arginine residue of β-NGF (Server and Shooter, 1976) it is also responsible for the removal of the two carboxy-terminal amino acids (Scott et al., 1983).

1.7 Relationship of γ-NGF and EGF-BP to other proteolytic enzymes

The studies described in the two previous sections, clearly show that for the two growth factors, NGF and EGF, the mechanism of their biosynthesis involves processing

into the biologically active form and individual packaging into stable complexes ready for export(Server and Shooter, 1976; Berger and Shooter, 1977; Frey et al., 1979). The two arginine-specific serine proteases involved in this mechanism, γ-NGF and EGF-BP, share very similar physical, chemical and immunological properties with each other and with other arginine-specific proteases found in the submaxillary gland (Bothwell et al., 1979). One of the characterised proteases in this class is β-NGF endopeptidase (Wilson and Shooter, 1979). γ-NGF, EGF-BP and β-NGF endopeptidase are all potent arginyl esteropeptidases as determined by their ability to cleave synthetic esters of arginine. They have very little general proteolytic activity and share similar inhibition properties (Wilson and Shooter, 1979). They all have molecular weights of around 26,000 d and very similar amino acid compositions as outlined in Table 1.1 (Thomas et al., 1981; Taylor et al., 1974; Wilson and Shooter, 1979). The strong conservation of their amino acid sequence (≃80% between γ-NGF and EGF-BP) and the demonstration that they share many common antigenic determinants clearly show that these three enzymes are members of a closely-related group of serine proteases (Thomas et al., 1981; Silverman, 1977; Bothwell et al., 1979). These similarities are further emphasised by the ability of β-NGF endopeptidase to form complexes, albeit of low stability, with either EGF or with the α and β-NGF subunits (Wilson and Shooter, 1979).

	γ-NGF	EGF-BP	β-NGF endo.
K	19	21	17
H	6	6	5
R	4	4	3
D+N	29	28	30
T	17	15	12
S	12	14	16
E+Q	14	21	17
P	16	15	19
G	20	23	20
A	12	14	12
C	10	10	11
V	12	11	12
M	7	3	4
I	9	7	12
L	23	24	27
Y	9	10	7
F	8	7	5
W	6	5	4

Table 1.1 Amino acid compositions of γ-NGF (Thomas et al., 1981), EGF-BP (Taylor et al., 1974) and β-NGF endopeptidase (Wilson and Shooter, 1979). Standard abbreviations are used for amino acids (Dayhoff, 1978).

In addition, all of these proteases were found to possess substantial kininogenase activity compared to their low general proteolytic activity. The kininogenase activity was measured by the ability to release vasoactive kinins from mouse plasma LMW-kininogen (Bothwell et al., 1979). This is a property normally attributed to the proteolytic enzyme glandular kallikrein which, like these enzymes, is a highly-specific arginyl esteropeptidase having little or no general proteolytic activity (Orstavik, 1980; Tschesche, et al., 1979). The β-NGF endopeptidase was found to be twenty-five fold more active as a kininogenase than either γ-NGF or EGF-BP (Bothwell et al., 1979). The rate of kinin release by β-NGF endopeptidase was comparable to that previously described for mouse submaxillary gland kallikrein (Porcelli et al., 1976). As essentially all the kininogenase activity of the submaxillary gland co-purifies with β-NGF endopeptidase, Bothwell et al., (1979) maintain that β-NGF endopeptidase is identical to mouse submaxillary gland kallikrein. The similarity of these proteases to glandular kallikreins from other sources further emphasises that γ-NGF, EGF-BP and β-NGF endopeptidase (kallikrein) belong to the same enzyme family as glandular kallikreins, for example, γ-NGF is 60% and 71% homologous with pig pancreatic kallikrein and rat submaxillary gland kallikrein respectively (Tschesche et al., 1979; Fiedler and Fritz, 1981; Lazure et al., 1981).

β-NGF endopeptidase was so named because of its ability to cleave an NH_2-terminal octapeptide from the β-NGF chain. Since β-NGF endopeptidase is the major form of salivary gland kallikrein, the biological relevance of its action on β-NGF is questionable. This is supported by the finding that the modified β-NGF chain retains its biological potency (Mobley et al., 1976) and no biological activity has yet been demonstrated for the released octapeptide (Angeletti et al., 1974).

Several reports (Schenkein et al., 1977; Wilson and Shooter, 1979) suggest that there are four additional kallikrein-like protease activities similar to γ-NGF, EGF-BP and β-NGF endopeptidase in the mouse submaxillary gland. These may include the serine protease γ-renin, which mimics the activity of the acid protease renin by cleaving angiotensin 1 from angiotensinogen in vitro. Amino acid sequence analysis of γ-renin (see Figure 3.3) suggests that it is very closely related to the kallikreins (Poe et al., 1983). The hypothesis that kallikrein (β-NGF endopeptidase), γ-NGF and EGF-BP are only a few members of a large group of enzymes functioning in the proteolytic processing of biologically active peptides is supported by the finding of other growth-promoting activities associated with arginyl esteropeptidases (Weimar and Haraguchi, 1975).

These include:-

(i) mesenchymal growth factor (MGF), which stimulates the growth of mesenchymal cells and causes a loss of differentiation in muscle and cartilage (Attardi et al., 1969).

(ii) thymocyte transforming factor (TTF), responsible for the transformation of small thymic lymphocytes into cells of the plasma series without cell division (Naughton et al., 1969).

(iii) factor two (F-2) which causes a powerful growth stimulation of mesenchymal cells (Weimar and Haraguchi, 1975).

A more detailed physical characterisation of these growth factors is necessary to determine if, like EGF and NGF, they are present as high molecular weight complexes comprising the growth factor and a specific kallikrein-like esteropeptidase activity.

1.8 Glandular kallikreins

As noted in the previous section, γ-NGF and EGF-BP are closely related to glandular kallikreins. This finding, as suggested in a recent review (Schachter, 1980), questions whether the many biological effects attributed to glandular kallikrein are the result of one protein or rather the cumulative actions of quite distinct but closely-related enzymes.

Kallikrein or kallikrein-like activities have been found in a wide variety of tissues in addition to the salivary gland (Werle, 1972). Recent studies using techniques of immunohistochemistry and immunoflourescence have defined the subcellular location of glandular kallikreins. Kallikrein has been localised to the tubular duct cells of the submaxillary gland (Orstavik et al., 1975; Simson et al., 1978), the acinar cells of the pancreas (Orstavik and Glenner, 1978), and the distal convoluted tubular cells of the kidney (Orstavik et al., 1979). Kallikrein-like antigens have been isolated from bovine mammary glands (Peeters et al., 1976) rat stomach (Uchida et al., 1980) and rat intestine (Moriwaki et al., 1980) as well as lung, brain and nerve tissue (Werle, 1972). Interestingly the prostate gland of the guinea pig, in addition to containing substantial kallikrein activity (Schachter et al., 1978), contains large amounts of NGF. This kallikrein activity may be due to γ-NGF.

The many proposed physiological roles for glandular kallikreins have recently been reviewed (Schachter, 1980). These include the regulation of local blood pressure, ion transport, cell proliferation, and prohormone processing. In summary, the properties of the characterised members of the kallikrein enzyme family, in particular that of strict substrate specificity, suggest that glandular kallikreins play a role in many biologically significant events involving the limited proteolysis of biologically active peptides.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and reagents

Adenosine triphosphate (ATP): Sigma

Ampicillin: Sigma

Acrylamide: Bio-Rad

Cesium chloride: Metallgesellschaft

Chloramphenicol: Sigma

Deoxynucleotide triphosphates (dGTP, dATP, dTTP, dCTP): Sigma

Dideoxynucleotide triphosphates (ddGTP, ddATP, ddTTP, ddCTP): P-L Biochemicals

α-^{32}P-cordycepin ($\simeq$3000 Ci/mmol): Amersham

α-^{32}P-dATP and α-^{32}P-dCTP ($\simeq$3000 Ci/mmol): Amersham and New England Nuclear

γ-^{32}P-ATP ($\simeq$7000 Ci/mmol): New England Nuclear

Dimethyl sulphate: Eastman-Kodak

Guanidine thiocyanate: Merck

Hydrazine: Pierce

Isopropyl-β-D-thiogalactopyranoside (IPTG): Sigma

N,N'-methylene bisacrylamide (Bis): Bio-Rad

Nitrocellulose 0.45 µm, 0.1 µm: Schleicher and Schüll

Phenol: Wako Chemical Industries

Piperidine: Sigma

Sephadex G-50 (medium): Pharmacia

Spectinomycin: Upjohn

TEMED (N,N,N'N'-tetramethylethylenediamine):
Sigma

Tetracycline: Sigma

Urea: Schwarz/Mann

5-Bromo-4-chloro-3-indolyl-β-D-
galactopyranoside (X-gal): Sigma

X-ray film (AR, or RP-5): Kodak

All other chemicals were analytical grade.

2.1.2 Enzymes

Unless otherwise indicated all enzymes were obtained from commercial sources and used in accordance with the manufacturer's specifications. The following enzymes sources were for:

Bal31 exonuclease: New England Biolabs

Calf intestinal alkaline phosphatase (CIAP):
Collaborative Research

Deoxyribonuclease I: Sigma

E.coli DNA polymerase I, Klenow fragment:
Boehringer-Mannheim and N.E. Biolabs

Mungbean nuclease: P-L Biochemicals

Proteinase K: Boehringer-Mannheim

Pronase: Boehringer-Mannheim

Restriction enzymes: N.E. Biolabs, BRL, and Boehringer-Mannheim

AMV reverse transcriptase: Life Sciences Inc.

Ribonuclease A: Sigma

S_1 nuclease: Sigma

Terminal transferase: Ratliffe Biochemicals

T4 DNA ligase: gift from N. Deacon

T4 polynucleotide kinase: N.E. Biolabs

2.1.3 Nucleic acids

HindIII synthetic "linkers" (5'-CCAAGCTTGG-3') and the synthetic M13 primer (5'-GTAAAACGACGGCCAGT-3') were a gift from Genetech Inc., San Francisco.

The synthetic kallikrein primer complementary to a portion of mGK-1 (5'-CAGGGGTTTTGGGCCAAAG-3') was a gift from J.Coghlan, Howard Florey Institute, Melbourne.

Oligo$(dT)_{12-18}$ and oligo(dT)-cellulose were obtained from Collaborative Research.

Carrier nucleic acids were either from Boehringer-Mannheim (E.coli tRNA) or Sigma (Herring sperm DNA).

Influenza RNA size markers were a gift from J. Blok.

Balb/c and C57BL/6 mouse genomic DNA were gifts from K. Reed.

Some of the Quackenbush mouse submaxillary gland total RNA used in this study was a gift from T. Lockett. Other RNA samples were purified from mouse submaxillary glands provided by M. Sleigh (Quackenbush is an outbreed strain of Swiss Webster mice).

2.1.4 Bacterial strains and DNA vectors

The following derivatives of E.coli K12 were used for transformation experiments and propagation of bacteriophages lambda and M13.

RR1: F^-, hsd S20 (r_B^-, m_B^-), ara-14 proA2, $lacY_1$, galK2, rpsL20 (Sm^r), xyl-5, mtl-1, supE44, λ^- (Bolivar et al., 1977).

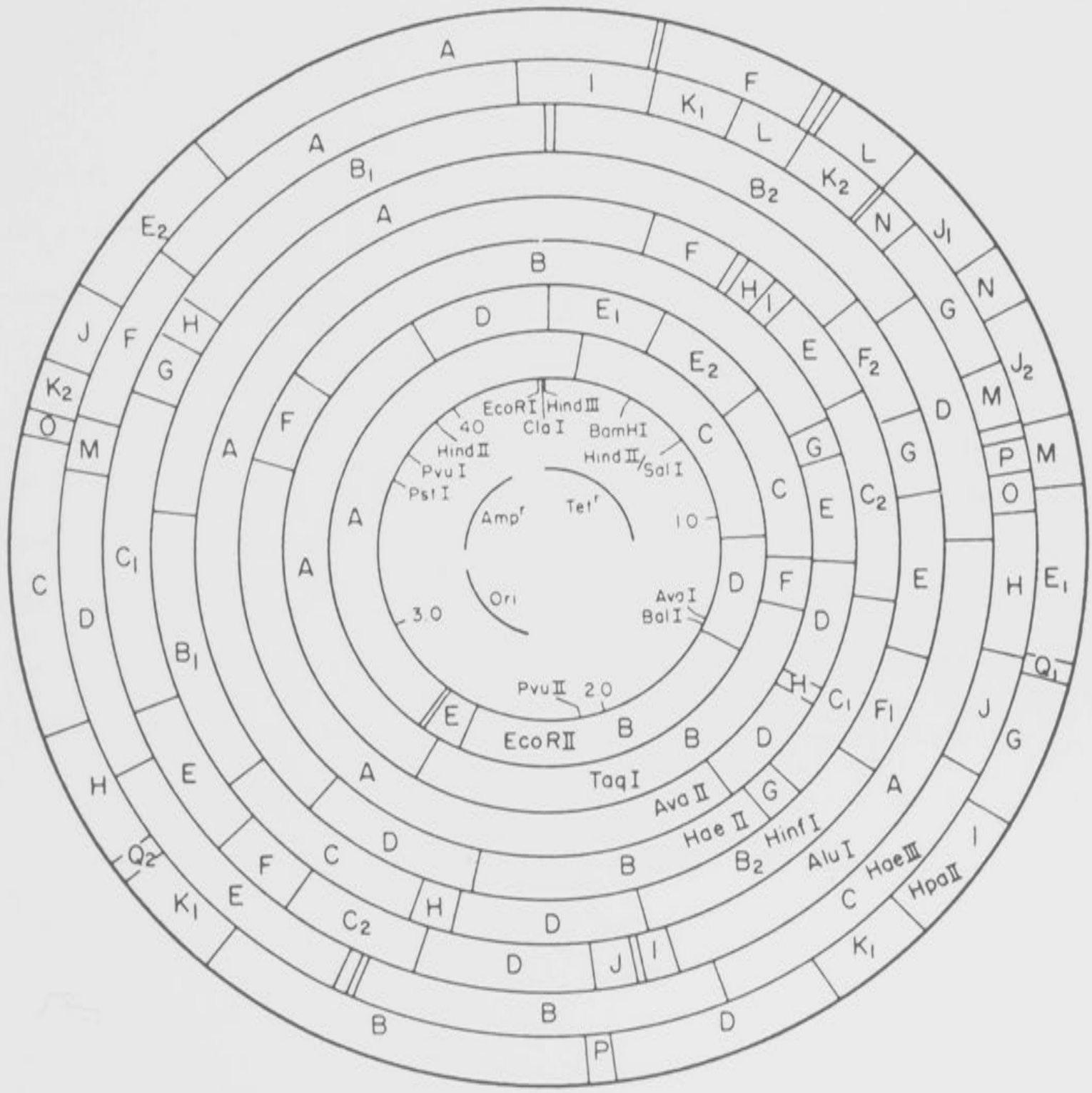
LE392: $F,^-$ hsd R514 (r_K^-, m_K^-), supE44, supF58, lacY1 or Δ(lac1ZY)6, galK2, galT22, metB1, trpR55, λ^- (Murray et al., 1977)

JM103: Δ(lac pro), thi, rpsL20(Sm^r) supE, endA, sbcB, $hsdR^-$, F traD36, proAB, $lacI^q$, ZΔM15 (Messing et al., 1981).

RR1 and JM103 cells were grown in Luria (L) broth (tryptone, 10g/l; yeast extract, 5g/l; sodium chloride, 5g/l). LE392 cells were grown in L-broth supplemented with 10 mM $MgCl_2$ and 0.2% maltose.

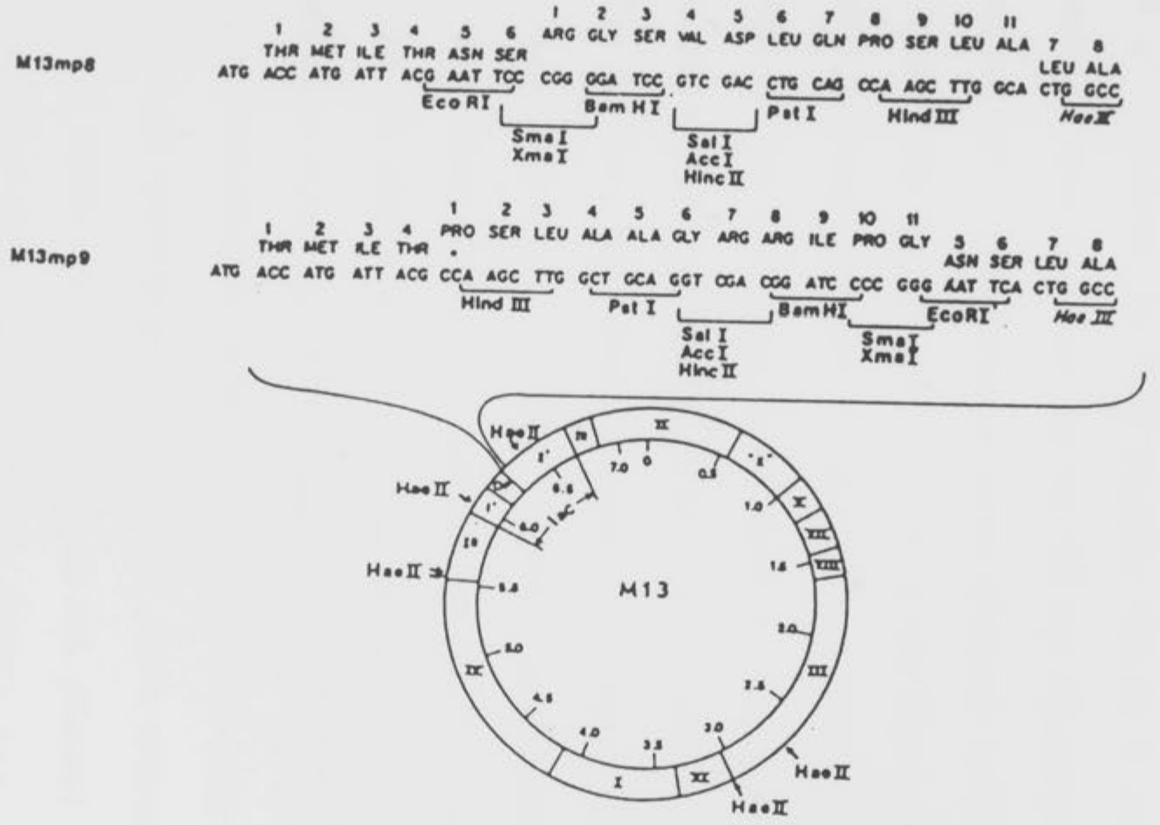
DNA fragments were cloned into the plasmid pBR322 (Ap^r, Tc^r) (Figure 2.1) and the M13 vectors mp8 and mp9 (Figure 2.2).

Figure 2.1 Restriction endonuclease map of the 4362 bp plasmid pBR322. The plasmid confers the phenotype Ap^r Tet^r. Inactivation of the Amp and Tet genes can be achieved by cloning fragments into the unique *Pst* I or *Bam*HI, *Hind*III and *Sal*I restriction sites, respectively. The position of these sites and the additional unique restriction sites for *Ava*I, *Pvu*II, *Cla*I and *Eco*RI are shown within the inner circle. The position and size of the fragments obtained by digestiion of pBR322 with any of the enzymes *Taq*I, *Ava*II, *Hae*II, *Hinf*I, *Alu*I, *Hae*III and *Hpa*II are shown on the outer circles; sizes are ordered alphabetically (Bolivar *et al*., 1977; Sutcliffe, 1978).


EcoRI
HindIII
ClaI
BamHI
HindII
PvuI
PstI
HindII/SalI
40
10
30
20
Amp^r
Tet^r
Ori
AvaI
BalI
PvuII
EcoRII
TaqI
AvaII
HaeII
HinfI
AluI
HaeIII
HpaII

pBR322

Figure 2.2 A genetic map of the single-stranded phage vectors M13 mp8 and M13 mp9. The HaeII - cleavage sites of the two different vectors have been mapped with respect to the genetic map (Messing et al., 1977). Map units are given in kb. The difference between the two molecules is shown in detail by the nucleotide sequence of the relevant region of the lac (β-galactosidase) gene. The recognition sites for various restriction endonucleases are underlined. The amino acid sequence corresponding to the polylinker fragment which was inserted into the original vector M13 mp2 is set up and numbered (reproduced in modified form from Messing and Vieira, 1982).


M13mp8
M13mp9
M13

2.2 Isolation of nucleic acids

2.2.1 Mammalian genomic DNA

High molecular weight DNA was prepared from liver essentially as described by Gross-Bellard et al., (1973). All manipulations were carried out at 4°C unless otherwise stated. 10 g of liver in NKM buffer (0.15 M NaCl; 5 mM KCl; 2 mM $MgCl_2$) was broken up by 10 to 12 strokes in a tissue homogenizer. The cells were washed twice in NKM buffer and lysed by the addition of 2 mM $MgCl_2$. An equal volume of TKM buffer (10 mM Tris-HCl, pH7.4; 0.2 M KCl; 2 mM $MgCl_2$) was added and the cell debris, containing intact nuclei, collected by centrifugation at 5000xg for 10 minutes. The pellet was resuspended in NKM buffer, washed twice in an equal volume of NKM buffer, and added dropwise to a slowly stirring solution of 10 mM NaCl, 10 mM Tris-HCl, pH8.0, 10 mM EDTA, 0.5% sodium dodecyl sulphate (SDS), 100 µg proteinase K/ml at 37°C over a 12 hour period. A further 100 µg proteinase K/ml of was added and the solution incubated overnight at 37°C. The solution was gently extracted with an equal volume of phenol saturated with TE (10 mM Tris-HCl, pH8.0; 1 mM EDTA) and chloroform (1:1). After centrifugation at 5000xg for 10 minutes, the solvent phase was removed and the aqueous phase extracted with an equal volume of chloroform in a similar fashion. The aqueous solution was dialysed against three changes of TE for 16 hours. The viscous solution was treated with

ribonuclease A (20 µg/ml, 37°C, 3 hours), phenol-chloroform extracted and dialysed against TE for a further 16 hours. The high molecular weight DNA (200-500 µg/ml) was stored at 4°C. Genomic DNA was prepared from tissue culture cells by the same procedure.

2.2.2 RNA isolation from tissue (Kaplan et al., 1979)

Mouse submaxillary glands were surgically removed from adult male Quackenbush mice and added to 5 volumes of extraction buffer (5M guanidine thiocyanate; 10 mM EDTA; 50 mM Tris-HCl, pH7.6; 5% v/v β-mercaptoethanol) and homogenized using an Ultra-Turrax blender. Sodium lauryl sarcosine was added to 4% (w/v) and the solution mixed by homogenization, followed by the addition 0.5 g of cesium chloride (CsCl) /ml of homogenate. 32 ml of this solution was carefully layered onto a 6.0 ml cushion of 5.7 M CsCl in a SW27 ultracentrifuge tube. After centrifugation (25,000 rpm, 16 hours), the supernatant was decanted and the pelleted RNA resuspended in TE, followed by the addition of 0.1 volume of 3 M NaOAc, pH6.0 and 2.5 volumes of ethanol. The RNA was kept at -20°C under ethanol until required.

2.2.3 Purification of poly(A)$^+$ mRNA (Aviv and Leder, 1972)

A column containing 0.5 g of preswollen oligo(dT)-cellulose was washed in 0.1 M NaOH, and equilibrated with

binding buffer (0.5 M NaCl; 10 mM Tris-HCl, pH8.0; 1 mM EDTA). Total RNA was resuspended in binding buffer at a concentration of 1 mg/ml, heated to 65°C for 5 minutes, cooled on ice and loaded onto the oligo(dT) column. The flow of RNA was monitored by A_{260} as measured by an ISCO ultra-violet (UV) spectrophotometer. After all the unbound RNA had passed through the column, the poly(A)$^+$ mRNA was eluted with TE buffer (10 mM Tris-HCl, pH8.0; 1 mM EDTA). The poly(A)$^+$ peak fraction was collected and further purified by an additional cycle of oligo(dT)-cellulose chromatography.

2.2.4 Supercoiled plasmid DNA (Clewell, 1972; Guerry et al., 1973)

RR1 cells containing the appropriate plasmid were grown overnight in L-broth containing antibiotics (either 20 µg Tetracycline/ml or 50 µg Ampicillin/ml). 10 ml of stationary phase cells were added to 1 litre of prewarmed L-broth and incubated at 37°C with shaking until the cells reached mid-log phase (OD_{600} = 0.4-0.5). At this point chloramphenicol (150 µg/ml) or spectinomycin (250 µg/ml) was added to the culture and incubation at 37°C was continued for 12-16 hours.

Cells were collected by centrifugation (5000 rpm, 5 min.), resuspended gently in 20 ml of ice-cold 25% sucrose in TE, and Lysozyme (5.0 ml, 5 mg/ml) and EDTA (10 ml, 0.25M, pH8.0) were added and the mixture was

incubated on ice for 10 minutes. 30 ml of Triton lysis mix (0.1% v/v Triton X-100; 62.5 mM EDTA, pH8.0; 20 mM Tris-HCl, pH8.0) was added and the mixture centrifuged at 18,000 rpm for 45 minutes in an SS-34 rotor. The supernatant was carefully decanted, treated with ribonuclease A (1 µg/ml, 30 minutes, 37°C) and subsequently with pronase (1 µg/ml, 30 minutes, 37°C). The solution was phenol-chloroform (1:1) extracted and the aqueous phase dialysed against three changes of 0.3 M NaCl in TE for 16 hours. 2.5 volumes of ethanol were added, the solution chilled at -20°C for 2 hours, and the precipitated DNA was pelleted by centrifugation (10,000 rpm, 15 minutes, SS-34).

Supercoiled plasmid DNA was prepared by CsCl/ethidium bromide density gradient centrifugation. The precipitated DNA was dissolved in a solution containing 2.8 ml of TE, 3.3 g of CsCl and 0.7 ml of EtBr (4 mg/ml) and centrifuged for 40 hours at 18°C (SW55 Ti, 45,000 rpm; Ti50, 45,000 rpm). The above volumes were scaled according to the amount of plasmid DNA to be purified. DNA in the gradient was visualized by UV light (350nm) and the lower supercoiled plasmid DNA band was removed by side-puncture with a 30 (gauge) needle. Ethidium bromide was removed by four sequential extractions with isobutanol saturated with TE. After an ether extraction to remove any remaining isobutanol, the plasmid DNA was dialysed against three changes of 0.3 M NaCl in TE for 16 hours. DNA was ethanol precipitated (2.5 volumes EtOH, -70°C, 1hour), resuspended

in water (concentration ≈ 1mg/ml) and stored frozen at -20°C.

2.2.5 M13 double-stranded(ds) replicative form (Messing et al., 1977)

One litre of L-broth was inoculated with 50 ml of stationary phase JM103 cells and 0.5 ml of M13-infected JM103 cells. Cells were harvested when they had reached stationary phase (37°C, 4-5 hours) and the M13 double-stranded replicative form was purified by the same procedure used for plasmids (2.2.4).

2.2.6 M13 single-stranded (ss) DNA (Heidecker et al., 1980)

M13 plaques were picked with an applicator stick, inoculated into 2 ml of L-broth and grown for 6 hours at 37°C. 1.5 ml of cells were centrifuged in an Eppendorf bench centrifuge for 60 sec. One ml of supernatant was aspirated and mixed with 300 µl of a 25% polyethyleneglycol 6000/2.5 M NaCl solution and incubated overnight at 4°C. Phage were pelleted by centrifugation (4 mins, Eppendorf) and resuspended in 100 µl of TE. The ssDNA (2-5 µg) was purified by phenol-chloroform (1:1) extraction and ethanol precipitation.

2.2.7 Bacteriophage lambda DNA (Maniatis et al., 1978)

Bacteriophage (10^5-10^6pfu) stored in PSB (0.01 M Tris-HCl, pH7.4; 0.1 M NaCl; 0.01 M $MgCl_2$; 0.05% gelatin) were adsorbed to 100 µl of a stationary phase culture of LE392 in 1 ml of LM broth at 37°C for 30 minutes. The infected culture was added to 40 ml of pre-warmed LM broth and incubated at 37°C with vigorous shaking, until lysis had occurred.

After cell debris had been removed by centrifugation (10,000 rpm, 15 min) the supernatant was treated with deoxyribonuclease I (10 µg/ml) and ribonuclease A (10 µg/ml) for 30 minutes at 37°C. Phage particles were precipitated by the addition of a fifth volume of 25% polyethyleneglycol 6000/2.5 M NaCl and incubation at 4°C overnight. The precipitated phage were pelleted by centrifugation (8000 rpm, 20 min.) were resuspended in 1 ml of ϕ.80 buffer (0.01 M Tris-HCl, pH7.4; 0.1 M NaCl) and incubated at 37°C with SDS (0.5%) and pronase (10 µg/ml) for 30 minutes. DNA was purified by phenol-chloroform (1:1) extraction and precipated by the addition of 0.1 volume of NaOAc, pH6.0 and 2.5 volumes of ethanol. After incubation at -20°C for 30 minutes the phage DNA was pelleted in an Eppendorf microfuge, washed in 95% ethanol and resuspended in water at a concentration of approximately 0.1 mg/ml. Phage DNA stocks were routinely stored at 4°C.

2.3 Enzymatic reactions

2.3.1 Restriction endonuclease digestion

DNA was restricted at concentrations of 50-200 µg/ml in HaeIII buffer (6.6 mM $MgCl_2$; 6.6 mM β-mercaptoethanol; 6.6 mM Tris-HCl, pH7.4). Salt was added to the reaction as specified by the manufacturer's conditions for each enzyme. Generally a two-fold excess of enzyme was used and the digest allowed to proceed for 2 hrs at 37°C (65°C for TaqI). Reactions were terminated either by heating at 65°C for 5 minutes or by the addition of EDTA (to 12.5 mM), followed by phenol-chloroform (1:1) extraction and ethanol precipitation (as above). The extent of the digestion was monitored by routine gel electrophoresis.

2.3.2 Bal31 digestion of DNA (Legerski et al., 1978)

The exonuclease Bal31 was used to generate a series of overlapping plasmid deletions for use in sequence analysis. Linearised plasmid DNA (25 µg) was incubated with 15 units of Bal31 at 30°C in a volume of 250 µl of reaction buffer (12 mM $CaCl_2$; 600 mM NaCl; 1 mM EDTA; 20 mM Tris-HCl, pH8.1). Aliquots were removed at 1 minute intervals, phenol-chloroform (1:1) extracted and ethanol precipitated. The extent of the reaction was monitored by agarose gel electrophoresis. Conditions were adjusted according to the extent of the deletion required.

2.3.3 Synthesis of high specific activity hybridisation probes (Taylor et al., 1976)

Radioactively-labelled DNA hybridisation probes were prepared by primed synthesis using the Klenow fragment of E.coli DNA polymerase I and random primers. The random primers of 8-12 nucleotides in length, were prepared by treating herring sperm DNA with DNase I and fractionation on a DEAE-Sephadex G-50 column. Plasmid DNA or isolated insert DNA (≃100-200ng) was denatured in the presence of 100 µg of primers by boiling for 2 minutes and cooling on ice for 1 minute. The reaction mixture was made up to 30 µl by the addition of cold deoxynucleotide triphosphates (dATP, dGTP and dTTP to 0.75 mM), 30 µCi of α^{32}-P-dCTP (≃3000 Ci/ mmol), reverse transcriptase buffer (final concentration: 50 mM Tris-HCl, pH8.3; 7 mM $MgCl_2$; 20 mM KCl; 20 mM β-mercaptoethanol) and 1 unit of E.coli DNA polymerase I (Klenow). After 30 minutes incubation at 37°C, the reaction was terminated by the addition of EDTA (to 12.5 mM) and phenol-chloroform (1:1) extraction. The aqueous phase was passed through a Sephadex G-50 column (1 x 6 cm, G-50 medium grade) equilibrated with 0.1M NaCl in TE. Approximately 400 µl fractions were collected from the column and their radioactivity monitored using a Geiger counter. The labelled DNA probe, which eluted in the void volume, was ethanol precipitated and its specific activity determined by Cerenkov counting. The specific activity obtained was between 10^7-10^8 cpm/µg DNA .

2.3.4 (a) Labelling 3'termini with α-^{32}P-dNTP's and reverse transcriptase (Roberts et al., 1979)

Restricted DNA (20 µg) was dissolved in a 40 µl reaction containing 50 mM Tris-HCl, pH8.3, 7 mM $MgCl_2$, 20 mM KCl, 10 mM β-mercaptoethanol, 30 µCi of the appropriate α-^{32}P-dNTP and 30 units of AMV RNA-dependent DNA polymerase (avian myeloblastosis virus reverse transcriptase) and incubated at 37°C for 30 minutes. The labelling reaction was terminated by the addition of EDTA (to 12.5 mM) and two successive phenol-chloroform (1:1) extractions followed by ethanol precipitation.

(b) Labelling 3'termini with α-^{32}P-cordycepin and terminal transferase (Tu and Cohen, 1980)

DNA samples which had been restricted with the enzymes PstI and SacI were labelled at their 3'termini with α-^{32}P-cordycepin. DNA (30 µg) was labelled in a 50 µl reaction volume containing 0.1 M potassium cacodylate, 2.5 mM Tris base, 1 mM $CoCl_2$, 2 mM dithiothreitol, pH6.9, 40 µCi of α-^{32}P-cordycepin and 12 units of terminal transferase. After 30 minutes incubation at 37°C, the reaction was terminated by the addition of EDTA (to 12.5 mM), phenol-chloroform (1:1) extraction and ethanol precipitation.

DNA labelled at the 3'termini by either method (a or b) was further restricted so that fragments only labelled at one end could be prepared for chemical degradation sequencing reactions (2.7.1).

2.3.5 Labelling 5'termini with γ-^{32}P-ATP and T4 polynucleotide kinase (Chaconas and van de Sande, 1980)

DNA fragments were treated with calf intestinal alkaline phosphatase (CIAP) in order to remove their 5' phosphate groups. The 5'termini were then labelled by incubation at 37°C for 30 minutes in a 20 µl mix containing 60 mM Tris-HCl, pH7.5, 8 mM $MgCl_2$, 10 mM dithiothreitol, 50 µCi of γ-^{32}P-ATP and 2 units of T4 polynucleotide kinase. The reaction was terminated by the addition of EDTA (to 12.5 mM) and phenol-chloroform (1:1) extraction, followed by ethanol precipitation.

Synthetic DNA fragments labelled in this fashion were purified by ion-exchange chromatography on a G-25 DEAE-Sephadex column.

2.3.6 Construction of double-stranded complementary DNA (ds cDNA)

Synthesis of ds cDNA was performed essentially as described by Ullrich et al., (1977). Single stranded (ss) cDNA was synthesised in a 150 µl reaction volume containing 20 µg poly(A)$^+$ mRNA, 50 mM Tris-HCl pH8.3, 20 mM KCl, 7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 20 µg of oligo $(dT)_{12-18}$, 1 mM dGTP, dATP, dTTP, 0.1 mM dCTP, 30 µCi of α-^{32}P-dCTP and 50 units of reverse AMV transcriptase. This reaction mixture was incubated at 42°C for 45 minutes and then terminated by the addition of EDTA (to 12.5 mM) and two sequential phenol-chloroform (1:1) extractions. The RNA-

DNA hybrid was purified from unincorporated deoxynucleotide triphosphates by G-50 chromatography (2.3.3). After ethanol precipitation, the RNA was hydrolysed at 70°C for 30 minutes in a 100 µl reaction containing freshly prepared 0.1 M NaOH and 1 mM EDTA. Hydrolysis was terminated by the addition of a tenth volume of 2M NaOAc, pH4.5 and 2.5 volumes of ethanol.

Second-strand synthesis was performed in a 100 µl reaction volume containing the ss cDNA, 50 mM Tris-HCl, pH8.3, 20 mM KCl, 7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mM each of dATP, dGTP, dTTP, 0.1 mM dCTP, 30 µCi of α-^{32}P-dCTP, and 25 units of AMV reverse transcriptase. After incubation at 42°C for 90 minutes, 2 units of E. coli DNA polymerase I (Klenow) were added and the reaction mixture incubated for a further 10 minutes at room temperature. The ds cDNA was purified by G-50 chromatography. The success of the second-strand synthesis was monitored by comparing its sensitivity to the single-strand-specific exonuclease S_1 with that of the single-stranded cDNA.

The hairpin loop of the ds cDNA was removed by incubating the DNA at 20°C for 10 minutes in a 50 µl reaction volume containing 0.3 M NaCl, 30 mM NaOAc, pH4.5, 4.5 mM $ZnSO_4$ and 100 units of S_1 nuclease. The reaction was terminated by the addition of a tenth volume of Tris base, pH10-11, EDTA (to 12.5 mM) and phenol-chloroform (1:1) extraction. ds cDNA was purified by G-50 Sephadex chromatography and ethanol precipitation.

The termini of the ds cDNA were repaired by incubation at 20°C for 10 minutes in a reaction volume containing 50 mM Tris-HCl, pH8.3, 7 mM $MgCl_2$, 20 mM KCl, 10 mM β-mercaptaethanol, and 1 µM dNTP's and 1 unit of E.coli DNA polymerase I (Klenow). The reaction was terminated by the addition of EDTA (to 12.5 mM), phenol-chloroform (1:1) extraction and ethanol precipitation.

Synthetic "linker" DNA, encoding the HindIII recognition sequence (5'-CCAAGCTTGG-3'), was ligated to the cDNA (2.3.8) before insertion into the HindIII site of plasmid pBR322 (Shine et al., 1977).

2.3.7 Dephosphorylation of vector DNA (Shine et al., 1977)

To increase the efficiency of vector-insert ligation, vector DNA was treated with alkaline phosphatase to remove the 5' phosphate groups. Vector DNA was treated with CIAP (0.1-0.2 units/µg DNA) in the presence of 100 mM Tris base, pH 10-11, 0.2% SDS for 1 hour at 37°C. The reaction was terminated by three sequential phenol-chloroform (1:1) extractions and ethanol precipitation. The precipitated DNA was washed twice with 70% ethanol and once with 95% ethanol to remove SDS.

2.3.8 Ligation of DNA molecules (Ullrich et al., 1977)

Cohesive-end ligations were performed at DNA concentrations of 20-50 µg/ml in the presence of

0.5 mM ATP, 8 mM $MgCl_2$, 10 mM dithiothretiol, 60 mM Tris-HCl, pH7.5 and 2 units of phage T4 DNA ligase. Ligation reactions were incubated overnight at 4°C. The extent of ligation was monitored by agarose gel electrophoresis, using an unligated control. Blunt-end ligations (including linker ligations) were performed as above but at higher DNA concentrations (100-200 µg/ml) and in the presence of 0.1 mM ATP.

2.4 Bacterial transformation and selection of recombinants

2.4.1 Transformation

Competent RR1 (for pBR322 plasmid recombinants) and JM103 (for M13 phage recombinants) cells were prepared and stored as 500 µl aliquots at -70°C (Morrison, 1979). An aliquot of the ligation mixture (2.3.7) was added to 200 µl of thawed competent cells, incubated on ice for 30 minutes and then heat shocked at 42°C for 90 seconds.

For RR1 transformations, the mix was left on ice for 30 minutes, inocubated into 5 ml of L-broth and incubated at 37°C for 1 hour. 100 µl aliquots of this culture were spread on selective media (containing 20 µg tetracycline/ml or 50 µg ampicillin/ml) and the plates incubated overnight at 37°C.

For JM103 transfections, subsequent to the heat shock, dilutions of the transfection mix were added to 200 µl of mid-log phase JM103 cells, 20 µl of 0.1M isopropyl-β-D-thiogalactopyranoside (IPTG), 10 µl of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (100 µg/ml in dimethyl formamide), and 2.5 ml of molten soft agar (0.7% agar in L-broth). This mixture was rapidly mixed and poured onto an L-plate and incubated overnight at 37°C. Recombinant phage were initially detected as those forming white or colourless plaques. Wild-type M13 form a blue plaque as they retain the ability to cleave the galactosyl residue from the synthetic galactosidase substrate X-gal.

2.4.2 Plasmid miniscreen

The presence of the plasmid pBR322 confers resistance to ampicillin and tetracycline (Ap^r,Tc^r) (Bolivar *et al.*, 1977). The insertion of a piece of DNA into restriction sites within these antibiotic resistance genes (e.g.: *Hind*III and *Bam*HI in Tc^r and *Pst*I in Ap^r) leads to a loss of this phenotype. Cells harbouring recombinant plasmids were identified initially by their Ap^sTc^r or Ap^rTc^s phenotypes. Colonies of the required phenotype were screened for the presence of a recombinant plasmid by the following procedure.

A portion of the colony was picked and incubated on ice for 30 minutes in the presence of lysozyme (1 µl, 5 µg/ml) EDTA (1 µl, 0.25 mM) ribonuclease A(1 µl, 1 µg/ml)

and TE (20 µl). 6 µl of sample dye (25% glycerol; 5% SDS; 0.2% bromophenol blue; 0.2% xylene cyanol FF) were added and the reaction mixture incubated at 65°C for 5-10 minutes. After vigorous vortexing for 1 minute the sample was electrophoresed on a 1% agarose gel (2.5.1). Supercoiled vector DNA was used as a size control.

2.4.3 Colony and plaque hybridisation (Grunstein and Wallis, 1979; Benton and Davis, 1977)

Plasmid gene banks, bacteriophage λ libraries and M13 plaques were screened for recombinants by hybridisation. After overnight growth, plate replicas were made on pre-cut 0.45 µm nitrocellulose filters (Schleicher and Schüll, BA85). Cells or phage present on the replica filters were lysed and the DNA denatured by placing the filter on Whatman 3 MM paper saturated with 0.5 M NaOH, 0.5 M NaCl, for 1-2 minutes. Filters were neutralized on 0.5 M Tris-HCl, pH7.2, 2 M NaCl, rinsed in 2 x SSC (0.3 M NaCl; 0.3 M trisodium citrate) and allowed to air dry before being baked at 80°C *in vacuo* for 2 hours. The filters were hybridised with the appropriate DNA probe as described in 2.6.1. After autoradiography, the nitrocellulose filters were aligned with both the X-ray film and the original plates to enable positively-hybridising colonies or plaques to be selected. Two to three successive rounds of purification and hybridisation were necessary to ensure plaque or colony homogeneity.

2.5 Gel electrophoresis of nucleic acids

2.5.1 Agarose gel electrophoresis

All agarose gels were electrophoresed in Tris-Acetate-EDTA (TAE) buffer (40 mM Tris; 5 mM NaOAc; 1 mM EDTA, pH7.8). 1-1.5% agarose gels were run in a vertical apparatus (165 mm x 165 mm x 1.5 mm) for the resolution of 0.7-23 kilobase (kb) DNA fragments. A fifth volume of SDS dye (20% sucrose; 5 mM EDTA; 1% SDS; 0.2% bromophenol blue) was added to the sample and electrophoresis was conducted at 150 V, 50 mA for 1-2 hours. The gel was stained with ethidium bromide, destained in water and the DNA visualized on an ultra-violet (UV) trans-illuminator (245 nm for analytical gels, 302 nm for preparative gels). Gels were photographed using Polaroid Type-107 film. This system was used for routine analytical purposes and for the preparation of DNA samples for chemical degradation sequencing (2.7.1).

For increased resolution of DNA fragments, particularly genomic DNA, a horizontal agarose gel system was employed. These gels (190 mm x 140 mm x 5 mm, or 190 mm x140 mm x 12 mm) were electrophoresed in TAE buffer at 40 V, 40 mA for 12-16 hours. Low-melting point agarose gels were used for preparative separation of DNA fragments required for cloning experiments or for use as hybridisation probes.

2.5.2 Acrylamide gel electrophoresis (Peacock and Dingham, 1967)

DNA in the size range of 700 base pairs (bp) to 20 bp was fractionated on 5-10% polyacrylamide gels. Vertical polyacrylamide gels were cast in a 165 mm x 165 mm x 1.5 mm frame and electrophoresed in Tris-borate-EDTA (TBE) buffer (50 mM Tris; 40 mM boric acid; 1 mM EDTA, pH8.3) at 200 V, 30 mA for 1-2 hours. Gels were prepared by diluting an appropriate amount of filtered 30% acrylamide/N,N'-methylene bisacrylamide (20:1) with TBE. Polymerisation was achieved by the addition of 0.5 ml of 10% (w/v) ammonium persulphate and 20 µl of N,N,N',N'-tetramethylethylenediamine (TEMED) to the 50 ml mixture.

Gels were stained and photographed as set out in 2.6.1.

2.5.3 Electro-elution of DNA

DNA was isolated from polyacrylamide or agarose gel slices by electro-elution. Gel slices were placed in dialysis tubing containing 0.25 x TBE (12.5 mM Tris; 10 mM boric acid; 0.25 mM EDTA, pH8.3) and placed in an electro-elution chamber containing 0.25 x TBE. The DNA was electrophoresed out of the gel slice into the buffer by applying a current of 10-15 mA for 1-2 hours. The DNA was then precipitated by the addition of 0.3 M NaOAc, pH6.0 and 2.5 volumes of ethanol.

2.5.4 DNA sequencing gel electrophoresis

The products of DNA sequencing reactions were resolved in thin denaturing (7.0 M urea) polyacrylamide gels. For 20% acrylamide gels, a 50 ml mixture containing 2.5 ml of 20 x TBE, 10 g of acrylamide, 0.5 g of N,N'-methylene bisacrylamide (Bis), 21 g of urea was filtered and degassed, 0.5 ml of 10% (w/v) ammonium persulphate, 20 µl TEMED added and poured into a 380 x 170 x 0.3 mM gel mould, and allowed to polymerize for several hours before use. The 20% gel was pre-run (15 W, 1500 V, 10 mA) in 1 x TBE (no urea) for 30 minutes before samples were loaded. After the sample had run the required length of time, the gel frame was removed and the sequencing gel transferred to a section of previously-used X-ray film. The gel was exposed to a sheet of X-ray film (Kodak X-omat RP-5 or AR film) at 70°C for 12-16 hours with intensifying screens.

10% (380 mm x 300 mm) and 5% (980 mm x 170 mm, 980 mm x 300 mm) gels with correspondingly-reduced amounts of acrylamide and Bis were run in 2 x TBE buffer. These gels were electrophoresed at between 1500-2000 volts, for varying lengths of time. The combination of running a sample on a 20% (2 hrs), 10% (3 hrs) and twice on a 5% gel (16 hrs and 36 hrs) allowed resolution of up to 500 nucleotides from the point of initiation of the DNA sequence.

2.6 Transfer of nucleic acids to nitrocellulose filters

2.6.1 Southern transfer (Southern, 1975)

DNA fractionated on agarose gels (2.5.1) was depurinated in 0.25M HCl (5-10 minutes), rinsed in water, and denatured in 0.5 M NaOH, 0.5 M NaCl (30-60 minutes). The gel was neutralised by washing in 0.5 M Tris-HCl pH7.2, 2 M NaCl (30-60 minutes). The time periods of these washes was varied according to the thickness of the agarose gel. The denatured DNA was transferred to a 0.45 µm nitrocellulose filter (Schleicher and Schüll, BA85) by blotting the gel in 20 x SSC (3 M NaCl, 3 M trisodium citrate) for 12-16 hours. The filter was washed in 2 x SSC, air-dried and baked at 80°C *in vacuo* for 2-3 hours.

Nitrocellulose filters were prehybridised in a solution of 3 x SSC, 0.1% SDS, denatured herring sperm DNA (10 µg/ml), *E.coli* tRNA (20µg/ml), 0.2% Ficoll, 0.2% bovine serum albumin Fraction V, 0.2% Polyvinylpyrrolidone and 0.05 M Hepes pH7.0 for at least 2 hours at 65°C. Hybridisation to a denatured DNA probe (2.3.3) was carried out in a minimal volume of hybridisation mix at 65°C for 16 hours. The filter was then washed in varying concentrations of SSC (0.1-2X) and at different temperatures (20-65°C) depending on the stringency of hybridisation required. Filters were then air-dried and

exposed for autoradiography at -70°C with intensifying screens for 1-7 days.

2.6.2 Northern transfer (Thomas, 1980)

One microgram of poly(A)$^+$ mRNA was denatured at 50°C for 1 hour in a 20 µl reaction mixture containing 2 µl of 100 mM Na phosphate buffer pH7.0, 10 µl deionised dimethyl sulphoxide and 4 µl of freshly-deionised 40% glyoxal. After incubation bromophenol blue and xylene cyanol FF were added to 0.05% and the sample was electrophoresed in a vertical 1.5% agarose gel run in 10 mM Na-phosphate buffer pH7.0. During electrophoresis the buffer was constantly recirculated to maintain a constant pH. After the sample had run the required distance, the gel was soaked in 6 x SSC and transferred to a 0.1 µm nitrocellulose filter by blotting overnight in 20 x SSC. The filter was baked *in vacuo* at 80°C for 2 hours and prehybridised for 16 hours at 42°C in 50% deionised formamide, 3 x SSC, 0.1% SDS, denatured herring sperm DNA (10 µg/ml), *E.coli* tRNA (20 µg/ml), 0.2% Ficoll, 0.2% bovine serum albumin Fraction V, 0.2% polyvinylpyrrolidone and 0.05 M Hepes, pH7.0. The filter was hybridised to a radioactive DNA probe at 42°C for 24-48 hours, washed and exposed to X-ray film as described in 2.6.1.

2.7 DNA sequencing reactions

2.7.1 Chemical degradation method (Maxam and Gilbert, 1980)

DNA fragments, labelled at one end, were prepared and purified as described in 2.3.4 and 2.5.3. The end-labelled DNA was subjected to chemical reactions specific for guanine (G), purine (P), pyrimidine (Y) or cytosine (C) bases.

The labelled DNA, dissolved in 30 µl, water was divided into four aliquots: G (5 µl), P (10 µl), Y (10 µl) and C (5 µl).

G reaction: 200µl of cacodylate buffer (50 mM Na cacodylate; 0.1 mM EDTA, pH8.1) and 1 µl of dimethyl sulphate were added to G and incubated at 21°C for 2.5 minutes. The reaction was terminated by the addition of 50 µl of G-stop mix (3 M NaOAc, pH6.0; 2.5 M β-mercaptoethanol; 1 mM EDTA; 0.1mg *E.coli* tRNA/ml), 750 µl of ice-cold ethanol and the DNA was precipitated at -70°C for 1 hour.

P reaction: 25 µl of formic acid was added to P and incubated at 21°C for 5 minutes, stopped by the addition of 200 µl of P-stop mix (0.3 M NaOAc, pH6.0; 0.1 mM EDTA; 25 µg *E.coli* tRNA /ml), 750 µl of ethanol and the DNA precipitated at -70°C for 1 hour.

Y and C reaction: 10 µl of H_2O was added to Y and 15 µl of 5 M NaCl to C, followed by the addition of 30 µl of hydrazine. After 5 minutes at 21°C, the reactions were stopped by the addition of respective stopmixes (Y-0.3 M NaCl; 0.1 mM EDTA, pH8.0; 25 µg E.coli tRNA/ml , C-0.1 mM EDTA, pH8.0; 25 µg E.coli tRNA/ml) and the DNA precipitated in ethanol at -70°C for 1 hour.

All four samples were precipitated, dissolved in 0.3 M NaOAc, pH6.0, reprecipitated, washed in ethanol and redissolved in 25 µl of 1 M piperidine. The strand-scission reaction was carried out at 90°C for 15 minutes. Piperidine was removed by evaporation in vacuo. The DNA was resuspended in 25 µl of water, dried in vacuo and dissolved in formamide loading dye (90% deionised formamide; 0.05% (w/v) bromophenol blue; 0.05% (w/v) xylene cyanol FF; 10 mM Tris-HCl, pH8.0; 0.1 mM EDTA). Samples were heated to 90°C for 1.5 minutes before loading onto denaturing polyacrylamide sequencing gels (2.5.4).

2.7.2 Chain termination method using bacteriophage M13 (Sanger et al., 1977; Schreier and Cortese, 1979)

Restriction fragments were cloned into appropriate M13 vectors (Figure 2.2) and single-stranded M13 template DNA was prepared as described in 2.2.6. The primer was a chemically-synthesised 17-mer (5' GTAAAACGACGGCCAGT-3')

complementary to codons 7 to 11 of the E. coli lac Z gene. 1 µl of 10 x RT buffer (50 mM Tris-HCl, pH8.3; 20 mM KCl; 7 mM $MgCl_2$; 10 mM β-mercaptoethanol) were mixed, heated to 65°C for 90 seconds and for a further 10 minutes at 37°C. Subsequently the DNA was diluted to 21 µl by the addition of 16 µl of 1 x RT buffer and 1 µl (1 unit) of E.coli DNA polymerase I (Klenow). 5 µl aliquots of this mix were distributed to four separate tubes containing 0.3 µCi α^{32}P-dATP (≈3000 Ci/ mmol) and 1 µl of the appropriate (G, A, T or C) reaction mix. The mixes contained the following ratios of dideoxy/deoxy nucleotide triphosphates.

	(µM)							
	dGTP	ddGTP	dATP	ddATP	dTTP	ddTTP	dCTP	ddCTP
G mix	5	250	5	-	50	-	50	-
A mix	50	-	5	250	50	-	50	-
T mix	50	-	5	-	5	250	50	-
C mix	50	-	5	-	50	-	5	250

After 10 minutes incubation at 21°C 1 µl of chase mix (1 mM dNTP's) was added to each tube and the reaction was allowed to proceed for a further 10 minutes at 20°C. The reaction was stopped by the addition of 8 µl of formamide loading dye (90% formamide; 0.05% (w/v) bromophenol blue; 0.05% (w/v) xylene cyanol FF; 10 mM Tris-HCl, pH8.0; 0.1 mM EDTA) and the DNA denatured by boiling for 2 minutes prior to loading on a DNA sequencing gel (2.5.4).

2.8 Mapping of RNA transcripts

2.8.1 Mung bean nuclease mapping (Sharp et al., 1980)

This technique was used to map exon-intron boundaries. Fragments (≃50 ng), labelled at the 5'termini or 3'termini of the strand complementary to mRNA (5'labelled Sau3A-PstI 400 bp fragment from pK-1-HSm.8; 3'labelled Sau3A-EcoRI 224 bp fragment pK-1-HSm.8; 5'labelled TaqI-EcoRI 900 bp fragment from pK-1-B2.6) were used as template in an annealing reaction with male mouse submaxillary gland poly(A)$^+$ mRNA (≃5 µg). The annealing reaction was performed at 66°C for 5 minutes and subsequently at 60°C for 12 hours, in a reaction volume of 10 µl containing 80% formamide, 0.4 M NaCl, 0.05 M PIPES, pH6.5, 1 mM EDTA. After incubation, the reaction mixture was diluted with 190 µl of ice-cold mung bean nuclease

buffer (0.29 M NaCl; 0.03 M NaOAc, pH4.6; 1 mM $ZnSO_4$). Mung bean nuclease (50 units) was added and the sample incubated at 37°C for 90 minutes. The protected DNA was phenol-chloroform (1:1) extracted, ethanol precipitated and resuspended in formamide loading dye (2.7.1). The size of the protected fragment was determined by electrophoresis on a 10% denaturing sequence gel (2.6.4). Maxam and Gilbert (2.7.1) sequencing reactions of a 900 bp TaqI-EcoRI fragment of known sequence were used as size markers.

2.8.2 Primer extension

The primer extension analysis was performed essentially as described by Hagenbuchle et al., (1980). A 58 bp TaqI-RsaI fragment was isolated from pK-1-B2.6 by polyacrylamide gel electrophoresis. Following electro-elution, the primer was dephosphorylated (2.3.7) and 5'labelled (2.3.5) using γ-^{32}P-ATP and T4 polynucleotide kinase. Five picomoles ($\simeq$150 ng) of labelled primer were hybridised with 30 µg of male mouse submaxillary gland poly(A)$^+$ mRNA in 50 µl of 80% formamide, 0.4M NaCl, 40 mM PIPES, pH6.5, 1 mM EDTA, for 1 hour at 50°C. The reaction was terminated by ethanol precipitation and dissolved up in 40 µl of RT buffer (2.3.4) containing 1 mM each of the four deoxynucleotide triphosphates. After the addition of 30 units of AMV reverse transcriptase the reaction mix was

incubated at 42°C for 1 hour. The resulting products were ethanol precipitated and treated with 0.1 M NaOH, before sizing on a 10% sequencing gel (2.6.4). Denatured 3' end-labelled HpaII fragments of pBR322 were used as size markers.

CHAPTER THREE

ISOLATION AND CHARACTERISATION OF A MOUSE GLANDULAR KALLIKREIN cDNA CLONE

3.1 Introduction

The molecular cloning of complementary DNA (cDNA) synthesised from a mRNA population is a well-established method for obtaining purified gene fragments for sequence analysis and use as hybridisation probes. During the isolation of cDNA clones which encode bioactive peptides synthesised in the male mouse submaxillary gland, cDNA sequences homologous to pig pancreatic kallikrein were identified (Richards et al., 1982). The submaxillary gland of the male mouse was chosen as a source of mRNA for these studies since it contains many different growth factors and other bioactive peptides as well as kallikrein activities. The latter comprise approximately 5% of the total soluble protein found in this tissue (Bothwell et al., 1979; Wilson and Shooter, 1979). Given the high level of kallikrein found in the male mouse submaxillary gland it is reasonable to assume that kallikrein mRNA is fairly abundant and probably comprises a similar 5% of the total mRNA. Based on this assumption, the technique of judging the predominance of a particular cDNA clone within the total cloned population by its extent of hybridisation to heterogeneous single-stranded cDNA was used in an attempt to isolate specific cDNA clones, including those encoding both NGF and EGF as well as kallikrein activites.

This chapter describes the isolation and characterisation of a cDNA clone coding for a member of the glandular kallikrein gene family. Although extremely

homologous in sequence with the three known mouse submaxillary gland kallikreins, γ-NGF, (Thomas et al., 1981) EGF-BP (Silverman, 1977) and γ-renin (Poe et al., 1983) it represents an uncharacterised member of this family.

The isolation of a kallikrein cDNA clone should enable the levels and type of kallikrein gene expression in different tissues to be determined, as well as providing a means of investigating the genomic organisation of the glandular kallikrein gene family.

3.2 Results

The work described in this thesis was carried out after the isolation of the cDNA clone pMK-1. Acknowledgement is made in the text to those involved in the stage preceding the characterisation of pMK-1.

3.2.1 cDNA library construction

Submaxillary glands were surgically removed by D. Catanzarro from adult male Quackenbush mice and the RNA extracted using guanidine thiocyanate and centrifugation through CsCl (2.2.2). Poly(A)-containing RNA was purified by two cycles of hybridisation to oligo(dT)-cellulose (2.2.3).

Double-stranded cDNA was constructed by R.I. Richards using sequential reverse transcriptase reactions with

submaxillary gland poly(A)$^+$ mRNA as template. The ends of the ds cDNA were blunt-ended by S_1 nuclease and repaired by E.coli DNA polymerase I (Klenow fragment) (2.3.6). Synthetic "linker" DNA encoding the HindIII recognition sequence (5'-CCAAGCTTGG-3') was ligated to the ds cDNA (2.3.8). After cleavage with the restriction endonuclease HindIII, the ds cDNA was fractionated on a 6% polyacrylamide gel. cDNA of 400 base pairs (bp) or greater was excised and electro-eluted from the gel (2.5.3). The ds cDNA was ligated into the HindIII site of the plasmid pBR322 and the resultant chimeric DNA was used to transform E.coli RR1 cells (2.3.8; 2.4.1). Five hundred recombinants were selected on the basis of their ampicillin-resistant, tetracycline-sensitive phenotype.

3.2.2 Selection of recombinants

The predominance of a particular cDNA clone within the total cloned population can be estimated by the extent of its hybridisation to single-stranded cDNA synthesised from total mRNA. If a cDNA clone contains a sequence corresponding to a mRNA species present at a level of 5% it will hybridise more strongly with the total heterogeneous ss cDNA than a clone representing an mRNA species present at less than 1% of the population. This approach was used to isolate cDNA clones representing mRNA species of approximately 1-5% abundance. 500 recombinants were screened by R.I. Richards for their ability to hybridise to

total ^{32}P-labelled male submaxillary gland ss cDNA by colony hybridisation (2.4.3). Weakly hybridising colonies adjudged to contain cDNA inserts corresponding to mRNA of 1-5% abundance were picked for further analysis. Both those which hybridised very strongly (presumably including the highly abundant amylase sequence) and those which did not hybridise significantly were discarded.

Plasmid DNA was prepared from these colonies (2.2.4) and examined for the presence and size of cloned DNA inserts by both agarose and polyacrylamide gel electrophoresis after cleavage with the restriction endonucleases HindIII and/or HaeIII. Plasmids containing inserts exceeding 400 bp in length were subjected to preliminary sequence analysis by R.I. Richards. For this purpose, HindIII-cleaved plasmid DNA was radioactively labelled with α-^{32}P-dATP using reverse transcriptase (2.3.4 a). The DNA was then cleaved with HaeIII and the labelled fragments subjected to sequence analysis by the chemical degradation method (2.7.1).

Possible reading frames derived from the 5' end of the mRNA sequences were compared by a computer program (SEQ) against all the peptide sequences of the Dayhoff atlas (Dayhoff, 1978).

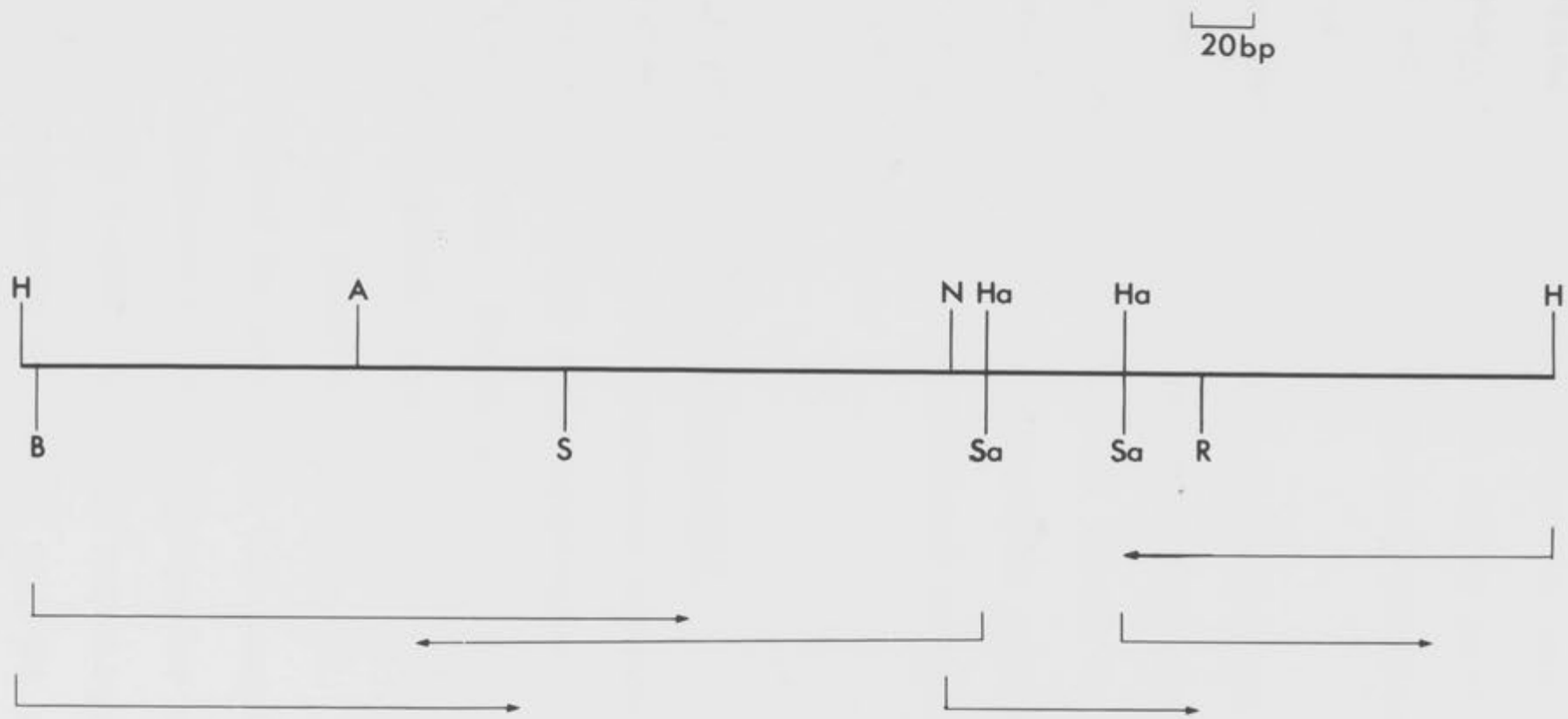
3.2.3 Identification and sequence analysis of pMK-1

The deduced amino acid sequence encoded from the 5' end of one recombinant plasmid (pMK-1) possessed 65%

homology with amino acids 95-124 of pig pancreatic kallikrein. This recombinant plasmid contained a 500 bp insert and was found to possess a poly(A) tract of 8 residues 19 bp downstream from the 5'-AATAAA-3' poly(A) addition signal sequence (Proudfoot and Brownlee, 1976). As this clone was assumed to code for a mouse kallikrein it was designated pMK-1. This initial identification was accomplished by R.I.Richards.

In order to confirm that pMK-1 contained a kallikrein-like sequence, the pMK-1 insert was completely sequenced. The clone was mapped for the presence of restriction sites suitable for further sequence analysis. Fig 3.1 shows the restriction map of pMK-1 deduced by a series of digests with the restriction enzymes AluI, BamHI, HaeIII, HindIII, HinfI, Sau3A, Sau96 and RsaI. The sequencing strategy used was based on the position of these restriction sites as outlined in Figure 3.1 and involved the cleavage of plasmid DNA with one of the enzymes, 3' end labelling of the resultant fragments with the appropriate α-^{32}P-dNTP and reverse transcriptase, followed by cleavage of the end-labelled fragments with another of the enzymes (2.3.4). Fragments labelled at only one end, and therefore suitable for sequence analysis, were purified by polyacrylamide gel electrophoresis and electro-elution (2.5.2, 2.5.3). The purified fragments were subjected to chemical degradation sequence reactions (2.7.1) and electrophoresed on 20% and 10% denaturing sequence gels

Figure 3.1 Restriction map of pMK-1 insert and sequencing strategy. The position of restriction sites within the insert are indicated by vertical lines. Restriction sites: H-HindIII, B-BamHI, A-AluI, S-Sau3A, N-HinfI, Ha-HaeIII, Sa-Sau96, R-RsaI. The arrows below the line indicate the length and direction of sequence obtained by the chemical degradation method (Maxam and Gilbert, 1980).


20bp
H
A
N
Ha
Ha
H
B
S
Sa
Sa
R

(2.5.4). Where possible the sequence was determined in both directions (Figure 3.1). To ensure that two or more restriction sites did not exist very close together, thereby resulting in possible errors in the ordering of sequencing data, all the cleavage sites used for sequence analysis were confirmed by sequence determination across the restriction sites.

The clone pMK-1 was found to contain an open reading frame coding for 149 amino acids terminating at the stop codon TGA, followed by a relatively short 3' untranslated region of 51 nucleotides. The complete nucleotide sequence of pMK-1 and its corresponding encoded amino acid sequence are presented in Figure 3.2.

3.2.4 Amino acid sequence comparisons

The predicted amino acid sequence encoded by pMK-1 was aligned with the known amino acid sequences of other kallikreins (Figure 3.3). These included mouse γ-NGF, mouse EGF-BP, mouse γ-renin, pig pancreatic kallikrein and also the sequence of pig pancreatic trypsin. The predicted peptide sequence encoded by pMK-1 is 57% homologous to that of pig pancreatic kallikrein, 76% homologous with the amino acid sequence of mouse γ-NGF, and 67% homologous to the partial amino acid sequences of mouse EGF-BP and γ-renin. The pMK-1 clone clearly contains only a portion of the coding potential (the COOH-terminal 149 amino acids) of a kallikrein-like peptide. By analogy to γ-NGF, one would

Figure 3.2 Nucleotide sequence of the 498 bp cDNA cloned in pMK-1. The predicted amino acid sequence encoded by the mRNA is shown above the nucleotide sequence. Nucleotides derived from synthetic oligonucleotide linkers are underlined. The poly (A) addition signal sequence 5'-AATAAA-3' in the 3'untranslated region is overlined (Proudfoot and Brownlee, 1976).

```
                Pro Glu Tyr Asp Tyr Ser Asn Asp Leu Met Leu Leu Arg Leu Ser Lys Pro Ala Asp Ile
TTG GAT         CCT GAG TAT GAC TAC AGC AAT GAC CTG ATG CTG CTC CGC CTC AGC AAG CCT GCT GAC ATC
                                          30                                        60

Thr Asp Val Val Lys Pro Ile Ala Leu Pro Thr Glu Glu Pro Lys Leu Gly Ser Thr Cys Leu Ala
ACA GAT GTT GTG AAG CCC ATC GCC CTG CCT ACT GAG GAG CCC AAG CTG GGG AGC ACA TGC CTA GCC
                        90                                      120

Ser Gly Trp Gly Ser Ile Thr Pro Thr Arg Trp Gln Lys Ser Asp Asp Leu Gln Cys Val Phe Ile
TCA GGC TGG GGC AGC ATT ACA CCC ACA AGA TGG CAA AAG TCA GAT GAT CTT CAG TGT GTG TTC ATC
                    150                                     180

Thr Leu Leu Pro Asn Glu Asn Cys Ala Lys Val Tyr Leu Gln Lys Val Thr Asp Val Met Leu Cys
ACG CTC CTC CCC AAT GAG AAC TGT GCC AAA GTC TAC CTA CAG AAA GTC ACA GAT GTC ATG CTG TGT
            210                                     240

Ala Gly Glu Met Gly Gly Gly Lys Asp Thr Cys Ala Gly Asp Ser Gly Gly Pro Leu Ile Cys Asp
GCA GGA GAG ATG GGT GGA GGC AAA GAC ACT TGT GCG GGT GAC TCC GGA GGC CCA CTG ATT TGT GAT
    270                                     300

Gly Ile Leu Gln Gly Thr Thr Ser Asn Gly Pro Glu Pro Cys Gly Lys Pro Gly Val Pro Ala Ile
GGT ATT CTC CAG GGA ACC ACA TCA AAT GGC CCT GAA CCA TGC GGT AAA CCT GGT GTA CCA GCC ATC
                                    360                                     390

Tyr Thr Asn Leu Ile Lys Phe Asn Ser Trp Ile Lys Asp Thr Met Met Lys Asn Ala
TAC ACC AAC CTT ATT AAG TTC AAC TCC TGG ATA AAA GAT ACT ATG ATG AAA AAT GCC TGA GTG TCA
                        420                                     450

CAT TGT CCC ATG TTC TTA ATA AAG CCC ACC ATG AAG CAA ATG (A)8 CCAA
                    480
```

Figure 3.3 Comparison of amino acid sequences of the serine proteases. Aligned, to maximise homology with the peptide encoded by pmK-1, are the partial amino acid sequences of γ-renin (Poe et al., 1983) and EGF-BP (Silverman, 1977) from mouse submaxillary gland, and the complete sequences of porcine pancreatic kallikrein (Tschesche et al., 1979), mouse submaxillary gland γ-NGF (Thomas et al., 1981), and porcine pancreatic trypsin (Hermodson et al., 1973). Residues common to the pMK-1 amino acid sequence are boxed. Residues common to all characterised serine proteases are indicated by a star (Young et al., 1978). Numbers refer to residues in the porcine pancreatic kallikrein sequence (Tschesche et al., 1979).

```
                  100                 120                 140
pMK.1     P E Y D Y S N D L M L L R L S K P A D I T D V V K P I A L P T E E P K L G S T C L A S G W G S I T P T
EGF-BP    [P E Y D Y] X [N D L M L]
γ-Renin
γ-NGF     L[E Y D Y S N D L M L L R L S K P A D I T D]T[V K P I]T[L P T E E P K L G S T C L A S G W G S I T P T]
P.P.Kall. D G K[D Y S]H[D L M L L R L]Q S[P A]K[I T D A V K]V L E[L P T]O[E P]E[L G S T C]E[A S G W G S I]E[P]G
Trypsin   - G N T L D[N D]I[M L]I K[L S]S[P A]T L N S R[V]A T V S[L P]R S C A A A[G]T E[C L]I[S G W G]N T K S S
                    *         *         *                       * *

                  160                 180
pMK.1     R - - W Q K S D D L Q C V F I T L L P N E N C A K V Y L Q K V T D V M L C A G E M G G G K D T C A G D
EGF-BP          F E N A K[D L Q C V]N L K[L L P N E]D[C]
γ-Renin   [W Q K]P[D D L Q C]M[F]T K[L L P N E N C]H[K]A H I L[K V T D]L[M L]X X I[E M]X E
γ-NGF     K - - F[Q]F T[D D L]Y[C V]N L K[L L P N E]D[C A K]A H I E[K V T D]A[M L C A G E M]D[G G K D T C]K[G D]
P.P.Kall. P D D F E F P[D]E I E[C V]Q L[T L L]Q[N]T F[C A]H A B P B[K V T]E S[M L C A G]Y L P[G G K D T C]M[G D]
Trypsin   G - - S S Y P S L[L Q C]L K A P V[L]S D S S[C]K S A[Y]P G Q I[T]G N[M]I[C]V[G]F L E[G G K D]S[C]Q[G D]
                                  *                   *                       * * *

                  200                 220                 240
pMK.1     S G G P L I C D G I L Q G T T S N G P E P C G K P G V P A I Y T N L I K F N S W I K D T M M K N A
EGF-BP
γ-Renin
γ-NGF     [S G G P L I C D G]V[L Q G]I[T S]W[G]H T[P C G]E[P]D M[P]G V[Y T]K[L]N[K F]T[S W I K D T M]A[K N]P
P.P.Kall. [S G G P L I C]N[G]M W[Q G]I[T S]W[G]H T[P C G]S A N K[P]S[I Y T]K[L I]F Y L D[W I]B B[T]I T Q[N]P
Trypsin   [S G G P]V V[C]N[G]Q[L Q G]I V[S]W[G]Y - G[C]A Q K N K[P]G V[Y T]K V C N Y V N[W I]Q Q[T]I A A[N]-
          * * * *               *   *       *
```

expect the pMK-1-encoded protein to be a total of approximately 233 amino acids in length (Thomas et al., 1981). The pMK-1-encoded protease fragment possesses many of the amino acids found to be common to all characterised serine proteases (Young et al., 1978), including the active site residues aspartic acid (residue 102) and serine (residue 197) (See Figure 3.3). The cysteine residue at position 157 is common to both trypsin and kallikrein-like activities and is also found in pMK-1 but not in any of the other serine proteases (Young et al., 1978; Tschesche et al., 1979).

γ-NGF, EGF-BP and γ-renin are members of a highly homologous group of enzymes which have been isolated from the mouse submaxillary gland and possess substantial kallikrein arginyl esteropeptidase activity (Bothwell et al., 1979). The extensive overall homology and conservation of functionally-important amino acids that the pMK-1 encoded protein shares with these proteases demonstrates that it belongs to the same enzyme family.

3.3 Discussion

In an attempt to characterise genes encoding the glandular kallikreins and several other bioactive peptides expressed in the male mouse submaxillary gland, a cDNA library was constructed using mRNA isolated from this tissue. Analysis of the cDNA library by hybridisation and

nucleic acid sequence analysis led to the identification of the recombinant plasmid pMK-1. This plasmid contains cDNA coding for a peptide sequence homologous with pig pancreatic kallikrein and three known mouse submaxillary gland kallikreins, γ-NGF, EGF-BP and γ-renin. Thus, the mRNA from which pMK-1 was derived appears to encode a previously uncharacterised member of the glandular kallikrein subgroup of the serine protease family. One possibility is that pMK-1 encodes a portion of β-NGF endopeptidase, a poorly characterised kallikrein in the salivary gland (Wilson and Shooter, 1979). However, a comparison of the amino acid composition of β-NGF endopeptidase (Wilson and Shooter, 1979; Table 1.1) with the sequence encoded by pMK-1 suggests that they are not identical. For example, β-NGF endopeptidase contains four Met and twelve Thr residues, whereas the pMK-1 sequence, which does not contain the complete coding sequence for the active peptide, already encodes one more of each of these residues. The protein encoded by pMK-1 is probably not the major form of mouse glandular kallikrein activity since Bothwell et al., (1979) have shown that β-NGF endopeptidase and most of the submaxillary gland kallikrein activity co-purify, suggesting that the same peptide is responsible for both activities.

The protein encoded by pMK-1 thus appears to be kallikrein-like but different from any of the presently known members of this enzyme group found in the

submaxillary gland. This suggests that there are at least five members (pMK-1, γ-NGF, EGF-BP, γ-renin, and β-NGF endopeptidase) of this enzyme family expressed in this tissue. This is in agreement with the data of Wilson and Shooter (1979) who found at least four enzyme activities in addition to EGF-BP, γ-NGF and β-NGF endopeptidase in the submaxillary gland. Given the known role of γ-NGF and EGF-BP (Frey et al., 1979), one may speculate that the pMK-1 encoded protein proteolytically modifies the precursor of a yet uncharacterised hormone or growth factor.

Trypsin-like esteropeptidases and kallikrein activities, as well as NGF and EGF, have been localised to the tubular duct cells of the mouse submaxillary gland (Ekfors and Hopsu-Havu, 1971; Simson et al., 1978; Schwab et al., 1976, and Gresik and Barka, 1977) by immunocytochemical and related techniques. Similar studies in the kidney have localised kallikrein activity to the distal convuluted tubules (Orstavik, et al., 1979). Hybridisation histochemistry experiments on mouse submaxillary gland and kidney tissue - two known sites of kallikrein synthesis, were performed by Jenny Penschow, Peter Hudson and John Coghlan at the Howard Florey Institute. Hybridisation experiments were conducted on 6 µm sections of tissue as described by Hudson et al., (1981b). Using ^{32}P-labelled pMK-1 insert (2.3.3), sequences homologous to pMK-1 were detected in the tubular

duct cells of the submaxillary gland (Fig 3.4) and the distal convoluted tubules of the kidney (Fig 3.5). These results have localised cross-reacting mRNA species to the same cells where kallikrein proteins are stored in both the submaxillary gland and kidney. These data reinforce the hypothesis that pMK-1 encodes a protein belonging to the glandular kallikrein group of enzymes, and shows that these cells are the site of kallikrein synthesis as well as protein storage.

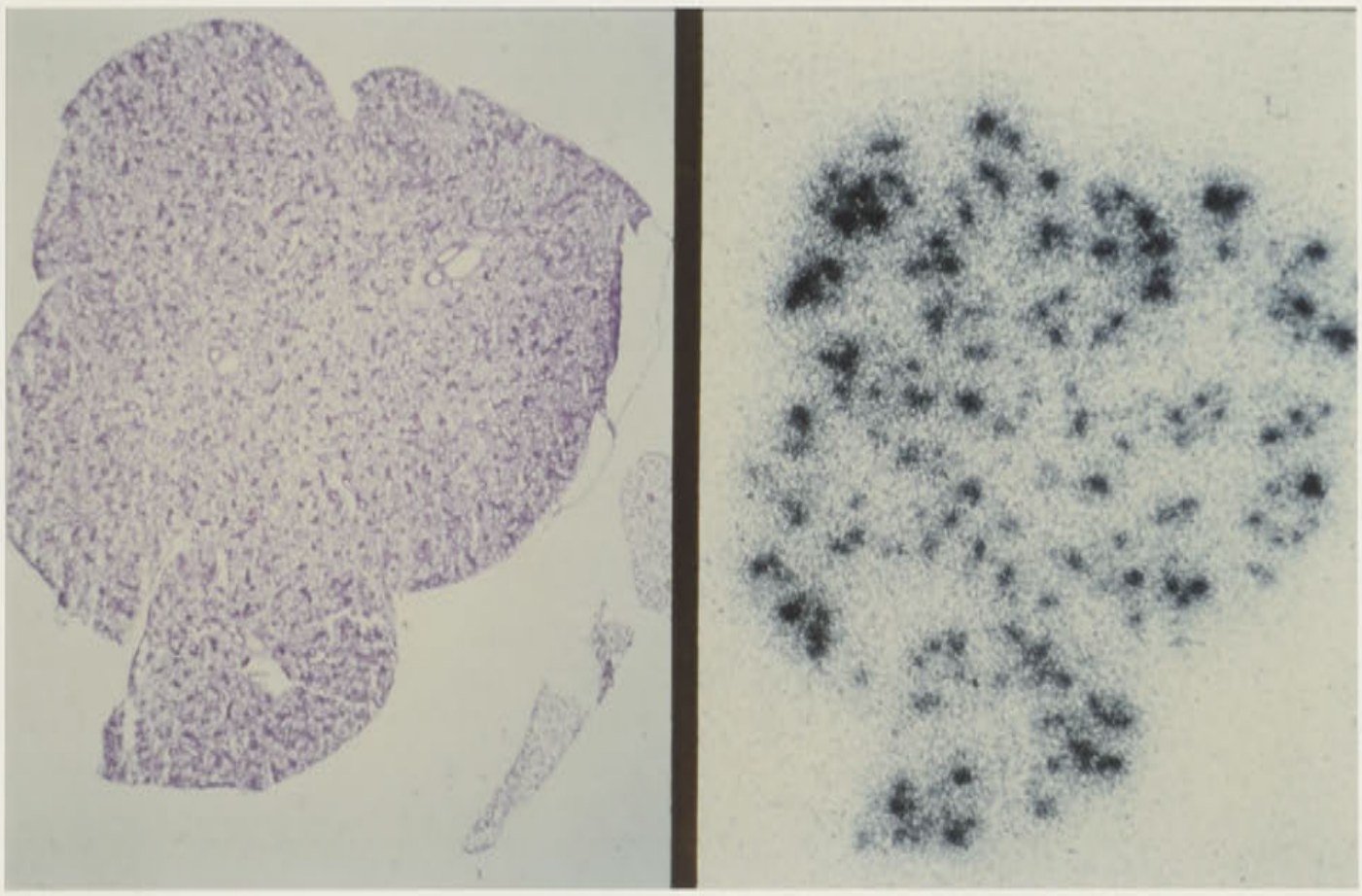
The isolation of a kallikrein cDNA clone provides a means of examining the level of synthesis of a variety of kallikreins in different tissues. The success of such experiments depends on the ability of pMK-1 to cross-react with other kallikrein mRNAs. The results presented in the following chapters demonstrate this cross-reactivity.

Figure 3.4 Hybridisation histochemistry used to detect the site of kallikrein synthesis at the mRNA level in mouse submaxillary gland. Tissue sections (6 µm) were cut on a cryostat, fixed and hybridised with a ^{32}P-labelled pMK-1 cDNA probe according to the method of Hudson (1981b).

(a) Tissue section of submaxillary gland stained with eosin and haematoxylin. Alongside is an autoradiograph of the next section hybridised with the pMK-1 probe (Magnification X 15).

(b) A thin section of submaxillary gland stained with haematoxylin and subsequently hybridised with the pMK-1 probe. The slide was coated with X-ray stripping film (AR-10, Kodak) and exposed for autoradiography. Silver grains are localised to the tubular duct cells of the submaxillary gland (Magnification X 140).

a



b

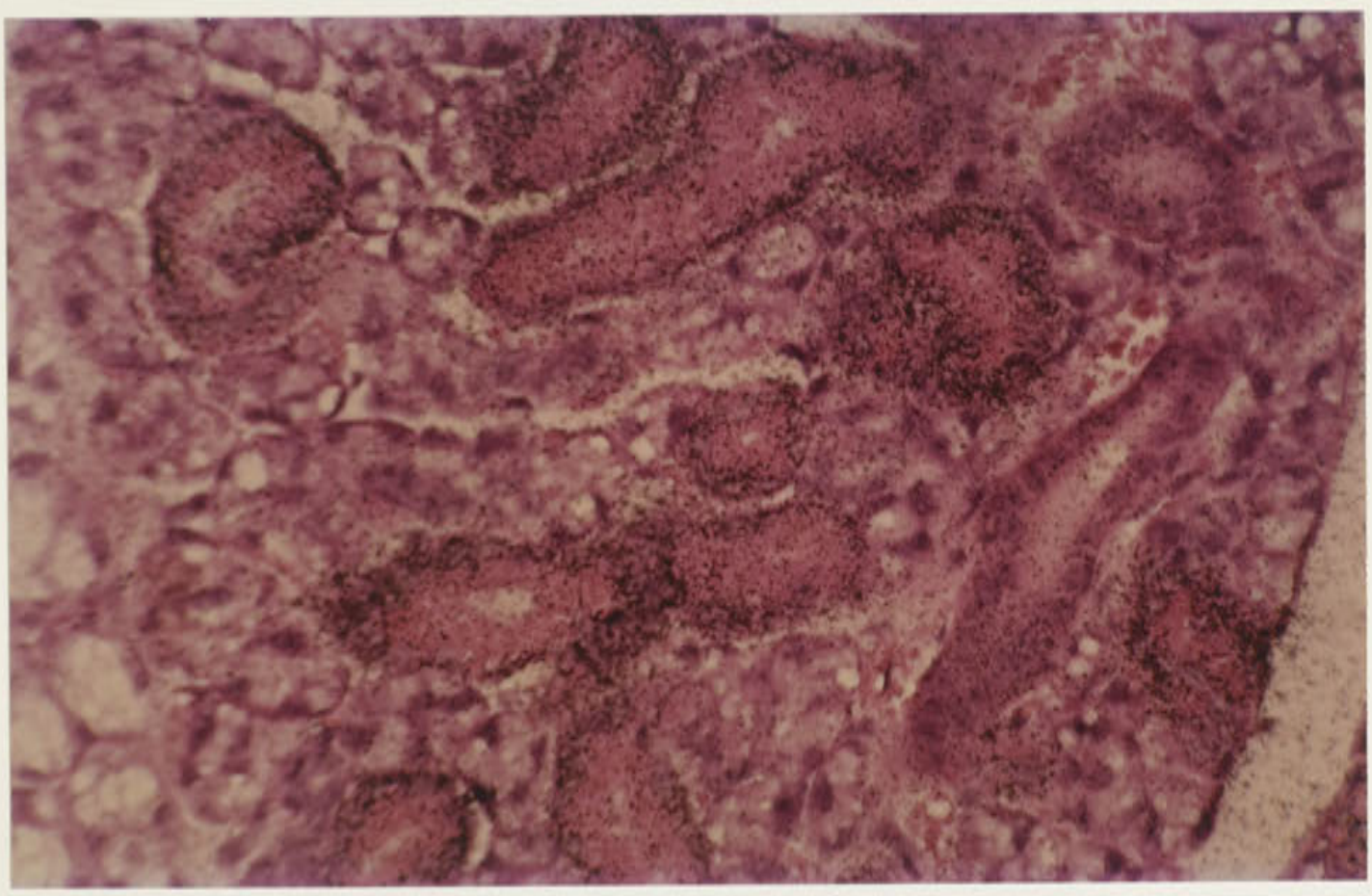
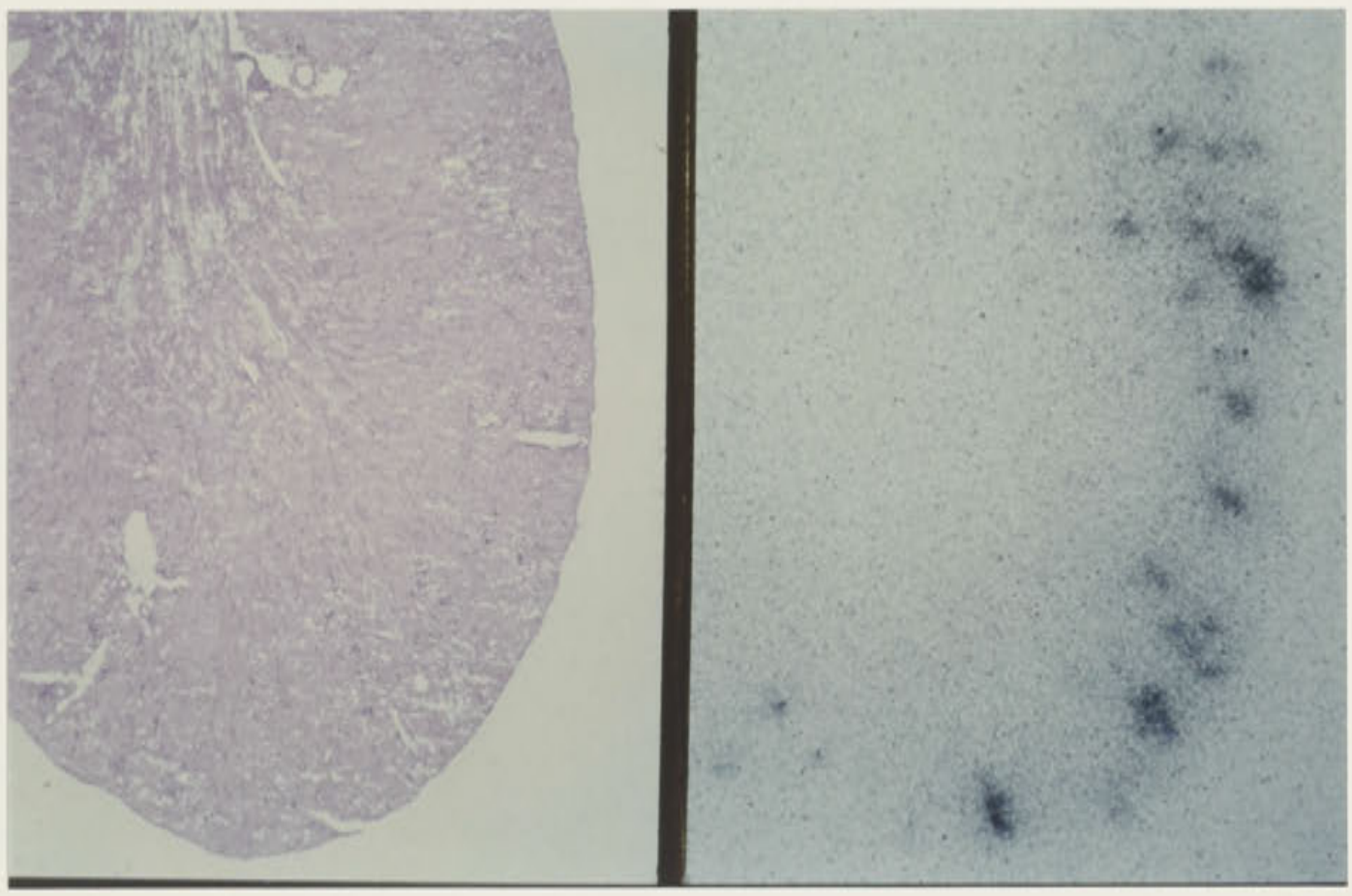


Figure 3.5 Hybridisation histochemistry used to detect the site of kallikrein synthesis at the mRNA level in mouse kidney. Tissue sections (6 µm) were hybridised with a ^{32}P-labelled pMK-1 cDNA probe according to the method of Hudson (1981b).

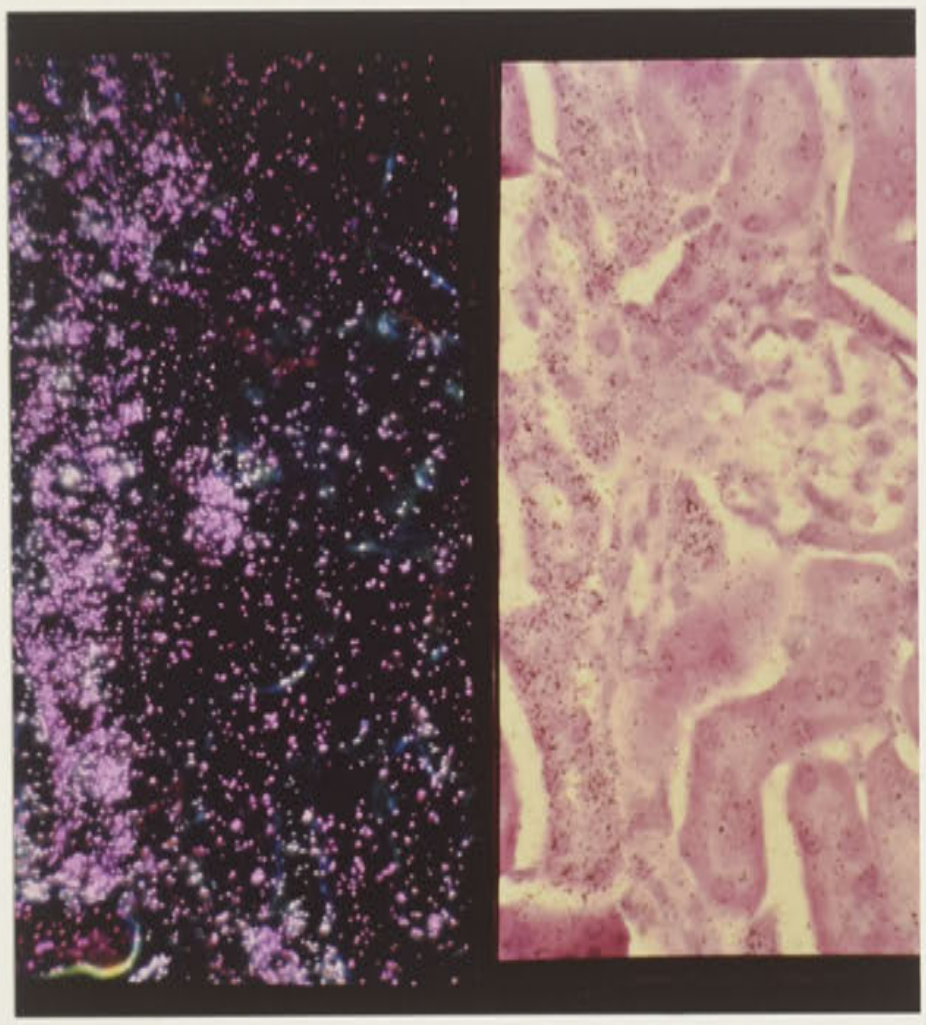
(a) Tissue section of kidney stained with eosin and haematoxylin. Alongside is an autoradiograph of the next section hybridised with the pMK-1 probe (Magnification X 15).

(b) A thin section of kidney stained with haematoxylin and subsequently hybridised with the pMK-1 probe. Sections were coated with X-ray stripping film and exposed for autoradiography. Shown is the autoradiograph viewed under dark-field microscopy (left) and light microscopy. The silver grains are localised to the distal convoluted tubules of the kidney. (Magnification X 160)

a



b



CHAPTER FOUR

GENOMIC ORGANISATION OF THE GLANDULAR KALLIKREIN GENE FAMILY

4.1 Introduction

The analysis of the complexity and organisation of the glandular kallikrein genes, as described in this chapter, was based on the ability of pMK-1 to cross-react with other members of this multi-gene family. The technique of genomic Southern blot hybridisation analysis can be used to gauge whether a particular sequence is unique, or whether it is homologous to other sequences in the genome. This approach was used to estimate the number of genes that would cross-react with pMK-1 at the DNA level. Since variation in the levels and type of kallikrein gene expression in the submaxillary gland has been reported in several mouse strains (Hiramatsu et al., 1981) the gene structure and copy number in different strains of mice was also compared by this technique.

A convenient method of mapping sequences to particular chromosomal locations is to make use of hybrid cell lines. The DNA from these cell lines, which contain an assortment of different chromosomes, can be analysed for the presence of a sequence by Southern blot analysis. The assignment of chromosomal location for any particular gene can then be made by correlation of the presence of a particular chromosome in the hybrid cell line and hybridisation to cloned sequences. The distribution of kallikrein genes within the mouse genome was investigated by using pMK-1 as a hybridisation probe on DNA isolated from Chinese hamster - mouse hybrid cell lines.

The results presented in this chapter demonstrate that in the mouse there is a family of highly homologous glandular kallikrein genes, all of which are located on chromosome seven. The hybridisation of pMK-1 to rat genomic DNA was used as a means of investigating the gene copy number and similarity of the kallikrein genes in another mammalian species.

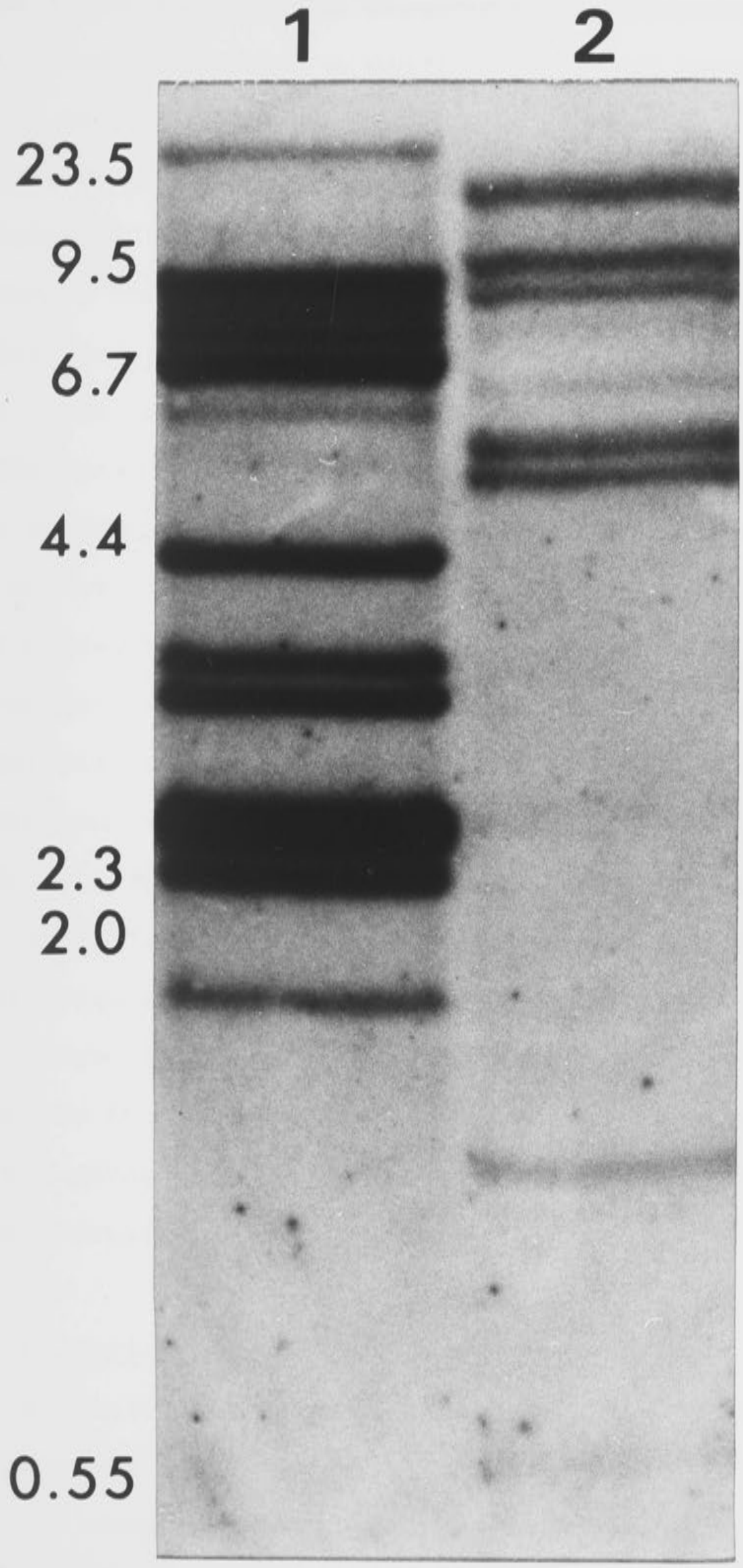
4.2 Results

4.2.1 Southern blot analysis of mouse and rat genomic DNA

High molecular weight genomic DNA was prepared as described in 2.2.1 and digested to completion with the restriction endonuclease EcoRI. The DNA (20 µg) was electrophoresed on a horizontal 1% agarose gel (2.5.1) and transferred by the method of Southern (2.6.1) to a nitrocellulose filter. The denatured DNA immobilized on the filter was hybridised with the random primed, ^{32}P -labelled 500 bp insert from pMK-1 under stringent conditions (see Legend Figure 4.1). Figure 4.1 shows the hybridisation pattern obtained when pMK-1 is used as a probe on EcoRI-cleaved Quackenbush mouse DNA (lane 1) and Sprague-Dawley rat DNA (lane 2). In both species multiple EcoRI fragments hybridise to pMK-1.

In mouse DNA there are at least 12 different sized restriction fragments which hybridise with pMK-1. The

Figure 4.1 Autoradiograph of a Southern blot of total genomic Quackenbush mouse DNA (Lane 1) and Sprague-Dawley rat DNA (Lane 2) digested with EcoRI and hybridised with the ^{32}P-labelled pMK-1 insert probe at 65°C in 3 x SSC buffer for 24 hours. After hybridisation, the nitrocellulose filter was washed in two changes of 2 x SSC at 65°C for 1 hour before autoradiography for 7 days at -70°C in the presence of an intensifying screen. Numbers alongside the autoradiograph represent the size in kilobases of markers derived from a HindIII digest of bacteriophage lambda c1857 DNA.


1
2
23.5
9.5
6.7
4.4
2.3
2.0
0.55

sizes of these restriction fragments were estimated, by comparison with those of a HindIII digest of bacteriophage λ DNA, to be 23.5, 8.8, 8.2, 7.4, 6.7, 6.4, 4.2, 3.4, 3.1, 2.4, 2.2 and 1.8 kb. Further studies on bacteriophage lambda genomic kallikrein clones (Chapter 6 and B. Evans, pers. comm.) have led to the identification of at least three distinct genes containing a 2.4 kb EcoRI fragment and another three genes with 6.7 kb EcoRI fragments. Therefore the differences in the intensity of hybridisation of certain restriction fragments (e.g: 2.4 and 6.7 kb) are due to the conservation of particular restriction sites around several different genes. The analysis of bacteriophage lambda genomic clones together with this data from Southern blot analysis suggests there are 25-30 distinct kallikrein genes (see Chapter 6).

The gene copy number in the rat appears to be similar to that of the mouse, with about ten EcoRI fragment sizes observed (Figure 4.1, Lane 2). It should be noted that the extent of hybridisation to the rat genome is significantly less than to the mouse genome, suggesting that the mouse kallikrein genes are more similar to each other than to their counterparts in the rat genome.

4.2.2 Comparison of different mouse strains

As variation in both the levels and type of kallikrein gene expression in the submaxillary gland have been reported in several mouse strains (Hiramatsu et al., 1981),

it was important to determine whether variation occurred at the DNA level in the kallikrein locus of different mice.

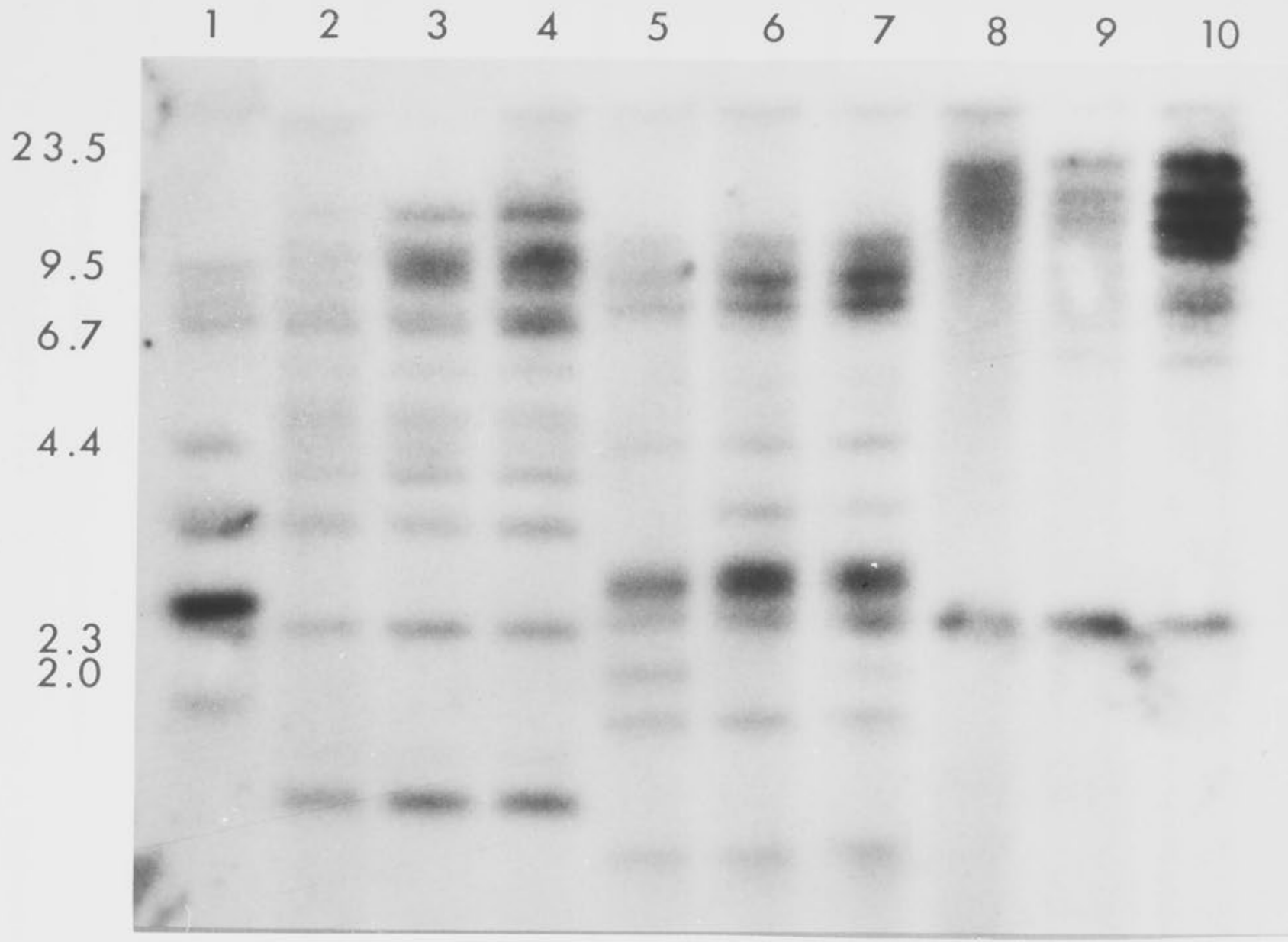
The pMK-1 hybridisation patterns of genomic DNA prepared from the Quackenbush, Balb/c and C57BL/6 strains of mice was compared after Southern blotting. DNA from these mice were cleaved with the restriction endonucleases EcoRI, SacI, BamHI and HindIII and hybridised to pMK-1 as described in 4.2.1. The hybridisation patterns observed with EcoRI- (data not shown) SacI- and HindIII-cleaved DNA (Figure 4.2) were indistinguishable from each other. However, there were slight differences in the hybridisation patterns observed with BamHI-cleaved DNA. The patterns observed for Balb/c DNA (lane 6, Figure 4.2) and Quackenbush DNA (lane 5, Figure 4.2) differ in a single uniquely hybridising band - 3.4 kb for Balb/c and 2.0 kb for Quackenbush. In each case this difference is probably due to a single restriction site polymorphism since the hybridisation pattern observed with C57BL/6 DNA (lane 7, Figure 4.2) includes both of these fragments.

The strong conservation of restriction sites in the kallikrein genes of these mice indicate both a similar structural organisation and kallikrein gene copy number in each of the three strains.

4.2.3 Chromosomal location of kallikrein sequences

The distribution of kallikrein genes within the mouse genome was investigated by using pMK-1 as a hybridisation

Figure 4.2 Southern blot analysis of Quackenbush, Balb/c and C57BL/6 mouse genomic DNA (ca. 7.5µg per lane) digested with HindIII (lanes 2-4 respectively); BamHI (lanes 5-7 respectively); SacI (lanes 8-10 respectively) and EcoRI (lane 1 - Quackenbush DNA) and hybridised with the ^{32}P-labelled pMK-1 probe. Hybridisation and washing conditions were as described in the legend to Figure 4.1. Numbers indicate the position of HindIII size markers.


1
2
3
4
5
6
7
8
9
10
23.5
9.5
6.7
4.4
2.3
2.0

probe on DNA isolated from a variety of Chinese hamster-mouse hybrid cell lines. These cell lines (supplied by D. Cox, UCSF) were constructed by fusing V-79 Chinese hamster cells with cells derived from two feral strains of mouse, Abruzzi and Cremona. Both of these mouse strains are homozygous for nine metacentric chromosomes derived by centric fusion of eighteen acrocentric autosomes. The Abruzzi mice have metacentric chromosomes in acrocentric combinations 1/7, 2/18, 3/8, 4/15, 5/17, 6/13, 9/16, 19/11, and 12/14 which differ from the acrocentric combinations of Cremona mice: 1/6, 2/8, 3/4, 5/15, 7/18, 9/14, 10/12, 11/13, and 16/17 (Cox et al., 1982). Somatic cell hybrids constructed from these mice strains and Chinese hamster cells contain differing numbers and combinations of mouse chromosomes and are thus ideal for mapping sequences to particular chromosomes.

All hybrid cell lines used to prepare DNA were karyotyped and/or analyzed by D. Cox for various characteristic enzymes in order to definitively assign the particular mouse chromosomes present in each hybrid line. DNA was prepared from each cell line as outlined in 2.2.1.

Figure 4.3 shows the hybridisation pattern observed with EcoRI-cleaved DNA from Quackenbush mice (lane 2), S-49 cells (mouse cell line, lane 3), Chinese hamster cells (lane 5) and III-16 cells (lane 4). III-16 is a Chinese hamster-mouse hybrid cell line containing mouse chromosomes in the following combination: 7, broken 1, 2/18, 3/8, 5/17,

6/13, 9/16, 12/14, 19 and X. It can be seen that the complete set of EcoRI fragments observed for Quackenbush mouse DNA is also present in the DNA prepared from the III-16 cell line. The pattern observed with the hybrid DNA differs in only one position from that of Quackenbush DNA (III-16 shows an additional 2.6 kb band). This is probably due to a single restriction site polymorphism - as previously observed between Balb/c, Quackenbush and C57BL/6 mice (4.2.2). These bands are not due to Chinese hamster sequences since under the hybridisation conditions used pMK-1 only weakly cross-reacts with Chinese hamster sequences (lane 5, Figure 4.3).

DNA was prepared from 12 different hybrid cell lines (Figure 4.6) as outlined in 2.2.1. These particular cell lines were selected as they possessed different combinations of the mouse chromosomes present in III-16 cells. The hybrid DNAs were cleaved with EcoRI and subjected to Southern blot hybridisation analysis (Figures 4.4, 4.5). It can be seen that when hybridisation to pMK-1 does occur, the complete set of EcoRI fragments is always present. Chromosome 7 is the only chromosome common to the DNAs which hybridised (Table 4.1) and therefore all the kallikrein genes that are detectable by hybridisation to pMK-1 are located on this chromosome.

Figure 4.3 Autoradiograph of a Southern blot of EcoRI-digested Quackenbush mouse DNA (lane 2), S-49 mouse cell line DNA (lane 3), III-16 Chinese hamster-mouse cell hybrid DNA (lane 4), and Chinese-hamster DNA (lane 5) hybridised with the pMK-1 probe. Conditions were as previously described (Figure 4.1). The markers (lanes 1 and 6) are ^{32}P-labelled λc1857 DNA digested with Hind III and are 23.5, 9.5, 6.7, 4.4, 2.3, 2.0 and 0.55 kb in size.

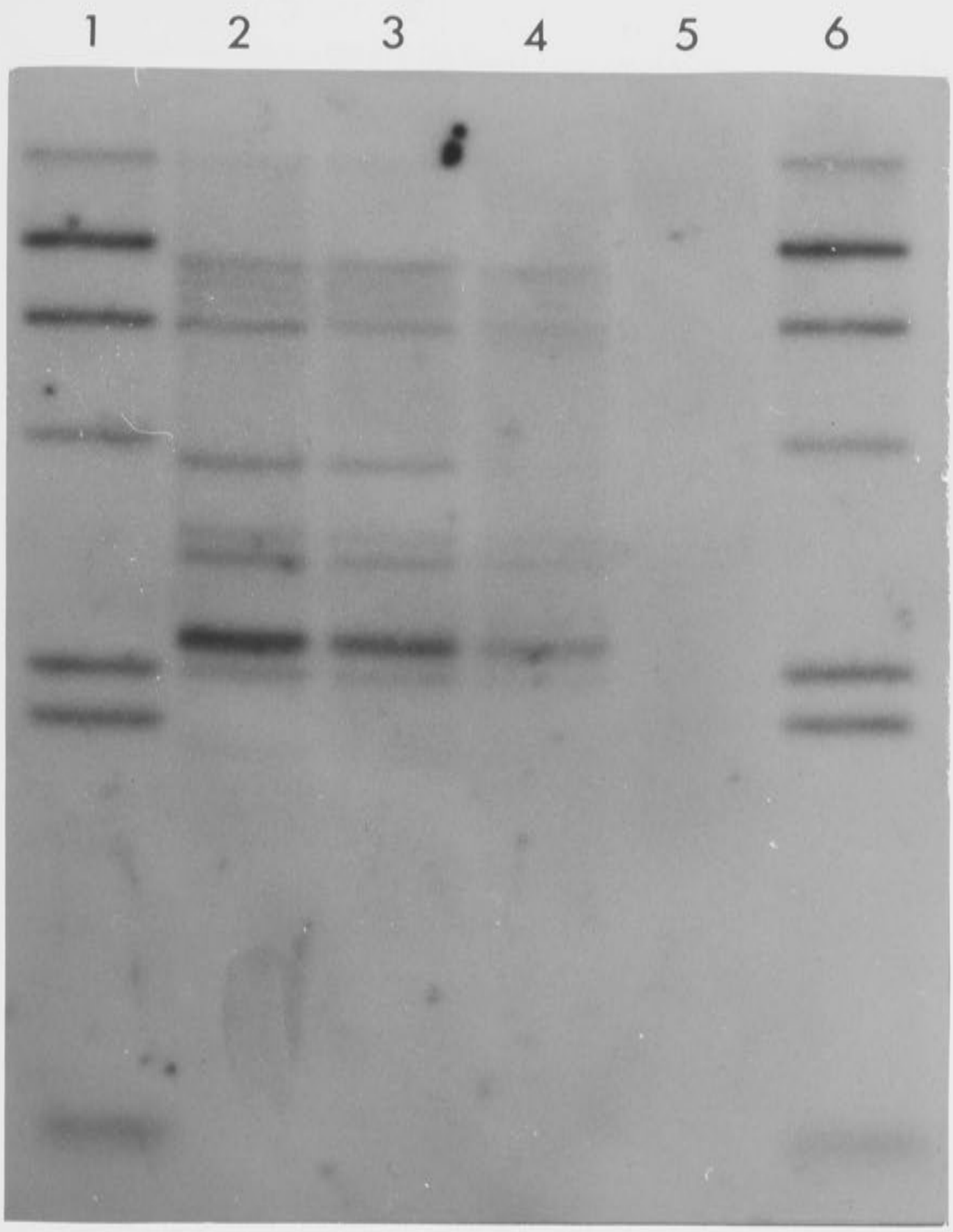

1
2
3
4
5
6

Figure 4.4 Southern blot analysis of DNA isolated from Chinese hamster-mouse cell hybrids. Approximately 7.5 µg of genomic DNA, prepared from cell lines: I-5 (lane 1), I-7B-4 (lane 2), I-3A-3 (lane 3), I-8-5 (lane 4), I-3A-2 (lane 6), III-23 (lane 7) and Chinese Hamster (lane 8), were digested to completion with EcoRI and fractionated on a 1% agarose gel, transferred to a nitrocellulose filter and hybridised with the pMK-1 probe (2.5.1, 2.6.1). The autoradiograph shown was obtained after exposure at -70°C for 7 days. The size of markers (lane 4) are 23.5, 9.5, 6.7, 4.4, 2.3 and 2.0 kb. The mouse chromosomes present in the hybrid cell lines are as outlined in Table 4.1.

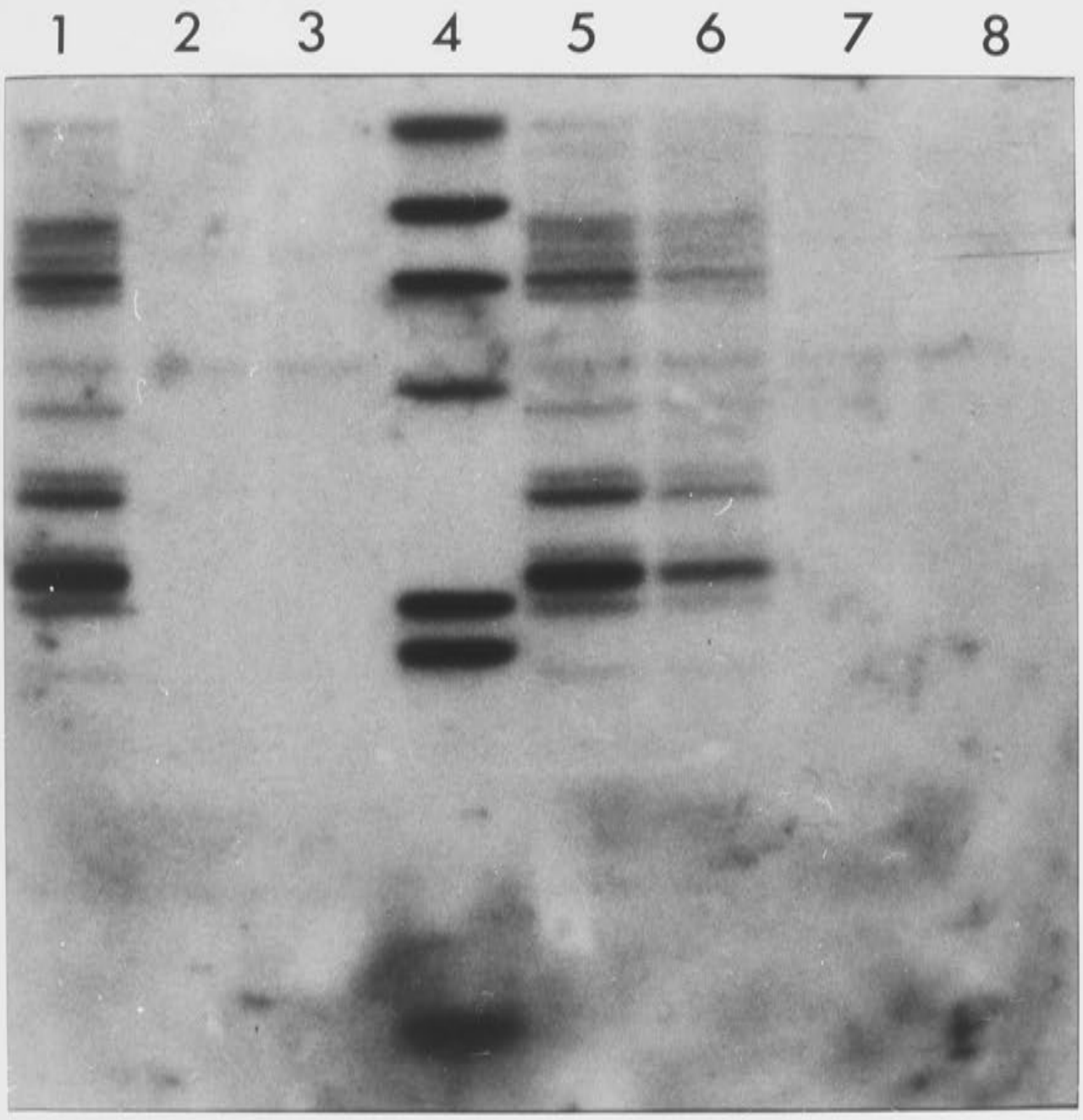

1
2
3
4
5
6
7
8

Figure 4.5 Autoradiograph of the Southern blot of DNA from hybrid cell lines III-12 (lane 2), I-8-2 (lane 3), III-13 (lane 4), CIV-12 (lane 5), III-14 (lane 6), III-16-1 (lane 7), which was hybridised with ^{32}P-labelled pMK-1 insert. Conditions were as described previously (Figure 4.1, 4.4). Lane 8 is a Quackenbush mouse DNA control and lane 1 shows the HindIII λ markers described in Figures 4.1, 4.4.

1 2 3 4 5 6 7 8

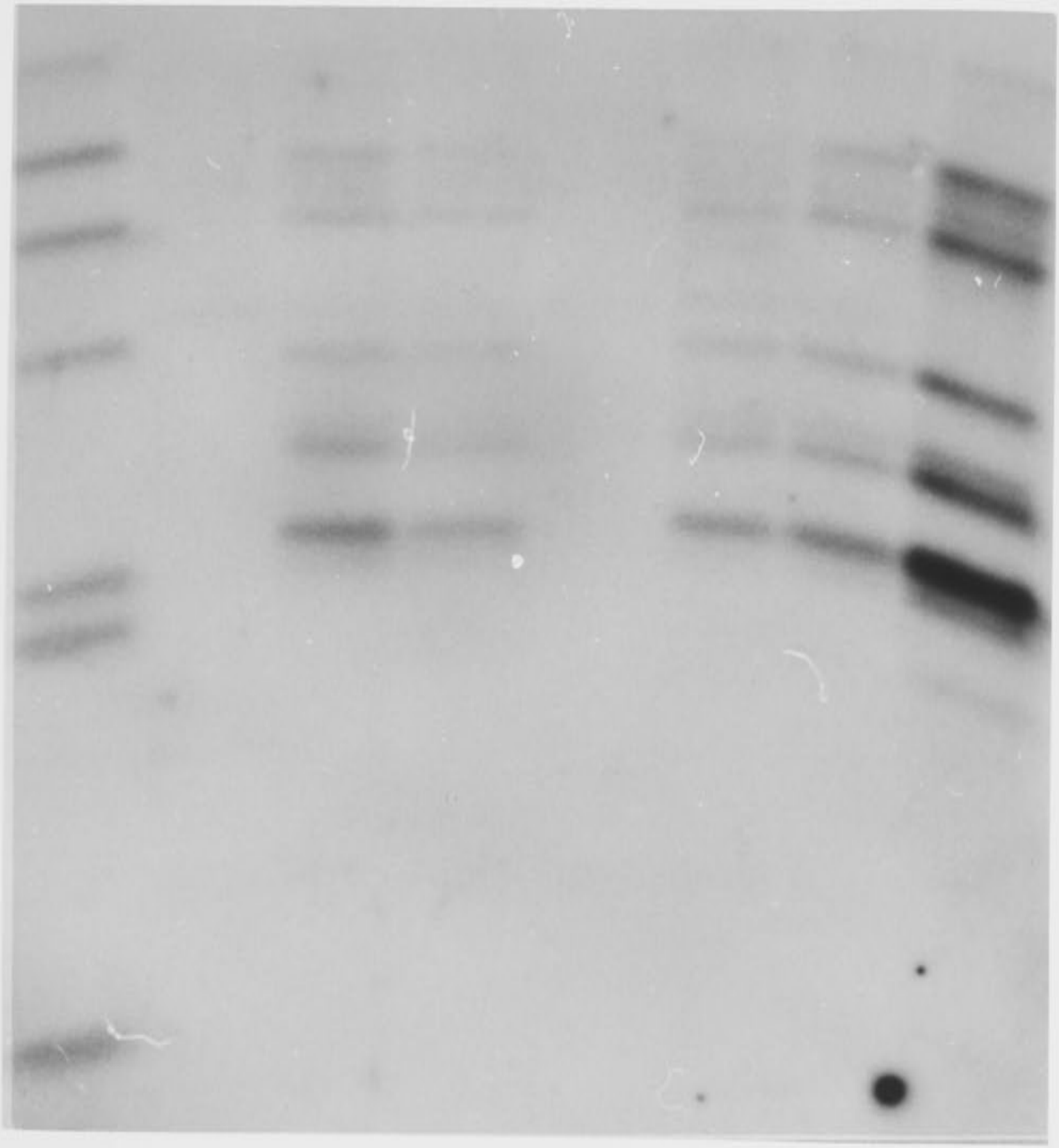


Table 4.1 Segregation of mouse chromosomes in Chinese hamster-mouse cell hybrids used for Southern blot analysis (Figures 4.4, 4.5). Hybrid cell lines which contained kallikrein sequences are designated with a star (*) and shown in the column marked 'hyb'. Only chromosome 7 is common to all these hybrid lines. (↑) This clone contained a rearranged 2/18 metacentric chromosome, involving a translocation of chromosomal material of unknown origin to chromosome 2 at band 2D, with deletion of the normal chromosome material distal to band 2D (Cox et al., 1982).

Clone	Mouse Chromosome																				Hyb.
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	X	
I-5*	+	+		+		+	+	+	+	+		+					+			+	+
I-7B-4	+	+								+				+	+			+	+		
I-3A-3	+	+		+					+				+		+			+	+		
I-8-5*		+		+		+	+			+		+			+				+		+
I-3A-2*		+		+			+	+				+	+		+	+		+	+		+
III-23		+[†]	+					+				+		+				+		+	
I-8-2*	+	+	+	+		+	+	+		+		+	+		+	+			+		+
III-13*	+						+		+							+				+	+
III-14*	+			+			+								+	+			+	+	+
III-12					+	+						+	+	+			+			+	
CIV-12		+						+		+		+				+	+			+	
III-16-1*		+	+		+		+	+	+			+		+		+	+	+	+		+

4.3 Discussion

The cross-reaction of pMK-1 sequences with a large number of differently sized restriction fragments in the mouse genome suggests that pMK-1 is a member of a highly homologous multi-gene family. This supports the hypothesis of Bothwell *et al.*, (1979) that γ-NGF and EGF-BP are members of a large group of enzymes sharing similar structural and physico-chemical properties. When compared with the data on kallikrein genomic clones presented in Chapter 6 these results show that there are some 25-30 members of this multi-gene family. Although based on only preliminary data, the hybridisation of pMK-1 to rat genomic DNA suggests that there are a similar number of kallikrein gene sequences in the rat. The extent of hybridisation to the rat genome is significantly less, suggesting that the mouse genes are more homologous to each other than to their rat counterparts. This is consistent with the results of immunological experiments which have shown that mouse γ-NGF, EGF-BP and β-NGF endopeptidase are more similar to each other than to rat submaxillary gland kallikrein(s) (Bothwell *et al.*, 1979).

There is a high degree of nucleotide sequence homology in the kallikrein gene loci of different strains of mice, as evidenced by the conservation of restriction sites and the extent of cross-hybridisation between Balb/c, Quackenbush and C57BL/6 mice. This data shows that the gene copy number and structural organisation of the

kallikrein gene locus is highly conserved in different mice strains. Therefore the variation in both the levels and type of kallikrein gene expression in different strains of mice (Hiramatsu et al., 1981) is probably due to differences in gene expression rather than polymorphism in gene number and/or type.

Many different loci encoding esterase activities in the mouse have been described (Womack, 1980). The only esterase activity that has previously been mapped to chromosome 7 by other techniques is the TAMase (tosyl-arginine-methylester) group of enzymes which, like the kallikrein genes, are arginyl esteropeptidases and are expressed in the mouse submaxillary gland. However, Skow (1978) has shown that the TAMases are both immunologically and electrophoretically distinct from γ-NGF, one of the members of the kallikrein gene family. Further chromosome mapping studies using cell lines containing deletions in portions of chromosome 7 are required to establish the relationship of the TAMase locus to that of the kallikrein genes.

The identification of a large number of DNA restriction fragments which cross-react with pMK-1 and the demonstration of their presence only in Chinese hamster-mouse hybrid cell lines containing the mouse chromosome 7, indicates that the members of the glandular kallikrein gene family are both highly homologous in primary structure and linked on this chromosome.

CHAPTER FIVE

ISOLATION AND SEQUENCE ANALYSIS OF A KALLIKREIN GENOMIC CLONE

5.1 Introduction

Elucidation of the structure and organisation of many eukaryotic genes has led to the discovery that a large number of protein-coding genes are interrupted by intervening sequences (introns). These studies have been made possible by the use of bacteriophage lambda gene libraries from which genomic clones containing regions homologous to cDNA sequences have been isolated (Maniatis et al., 1978). An examination of the molecular anatomy of eukaryotic genes has made it possible to develop theories on their evolution, possible mechanisms of RNA splicing and the nature of sequences involved in gene expression (Breathnach and Chambon, 1981).

To determine in more detail the organisation of kallikrein genes and to elucidate the molecular anatomy of a member of this multi-gene family, genomic clones homologous to the cDNA clone pMK-1 were isolated from a library of cloned mouse DNA sequences in bacteriophage lambda. This has allowed the exon-intron arrangement of a kallikrein gene to be determined and may provide some insights into the evolution of the multi-gene family as well as providing information on the complete structure of a preprokallikrein protease. The demonstration that all the kallikrein gene sequences are present on chromosome 7 suggests that these genes may be closely linked, as previously described for most mammalian multi-gene families, for example the growth hormone family (Kidd and

Saunders, 1982) and the α and β globins (Lauer et al., 1980; Efstratiadis et al., 1980). The work discussed in this chapter deals with the isolation and sequence analysis of a genomic clone (λMSP-1) which was found to contain two linked kallikrein genes, mGK-1 and mGK-2. In addition, the gene mGK-1 was shown to be transcriptionally active in the male mouse submaxillary gland.

5.2 Results

5.2.1 Isolation of genomic clones

A bacteriophage lambda genomic library of Quackenbush mouse DNA was constructed by ligating a partial MboI digest of high molecular weight DNA (10-20 kb) into BamHI-cleaved Charon 28 DNA (Liu et al., 1980). The recombinant DNA was packaged in vitro by R.I. Richards according to the method of Hohn (1979) and 30,000 clones isolated.

Bacteriophage from the cloned mouse DNA library were plated out at high density and screened with the pMK-1 hybridisation probe (2.4.3). Two independent lambda clones, λMSP-1 and λMSP-2, were isolated in pure form after three rounds of plaque purification (Benton and Davis, 1979). One of these clones, λMSP-1, was chosen for detailed analysis. Studies on the other clone, λMSP-2 will be discussed in Chapter 6.

5.2.2 Restriction analysis of λMSP-1

λMSP-1 phage DNA was purified as described in 2.2.7 and cleaved in either single or double digestions with the restriction endonucleases BamHI, EcoRI, SacI, HindIII, SmaI, and BglII. The digests were then fractionated on a 1% agarose gel and the DNA transferred to a nitrocellulose filter. The filter was hybridised to a ^{32}P-labelled pMK-1 probe and exposed for autoradiography as described 2.6.1. The autoradiograph of one of these gels is shown in Figure 5.1.

The data from these restriction digestions, subsequently confirmed by sequence analysis, enabled the construction of a physical map of the genomic clone λMSP-1 as shown in Figure 5.2. This clone comprises 13.5 kb of genomic DNA and contains two closely linked kallikrein genes which are transcribed in the same direction. The first gene (mGK-1) is complete in this clone (see below) and consists of five exons and four introns, spanning 4.5 kb. The 3' end of another gene (mGK-2) is located 3.7 kb upstream from the 5' end of mGK-1. In λMSP-1 no other kallikrein gene sequences could be detected downstream from mGK-1.

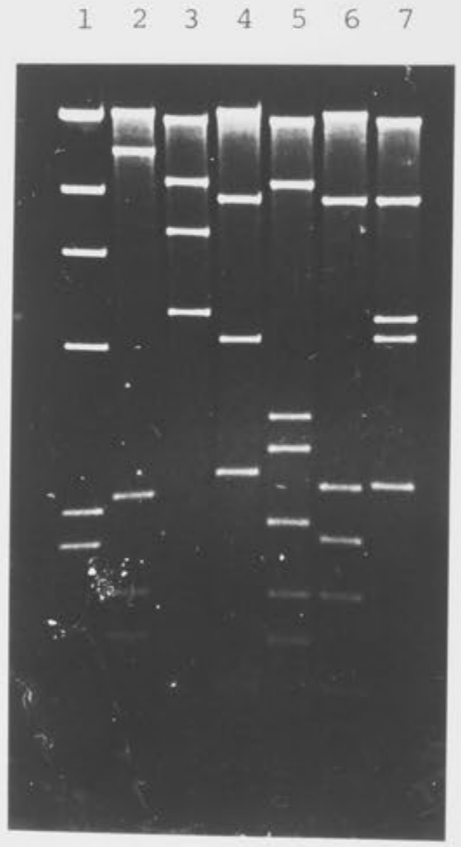
The sequences present in the cDNA clone pMK-1 were found to hybridise with corresponding sequences in mGK-1 and mGK-2, although they were interrupted by the presence of two intervening sequences (Figure 5.1, 5.2). As pMK-1 is only a partial cDNA clone and contains the coding

Figure 5.1 Restriction and Southern blot analysis of λMSP-1.

(a) Ethidium bromide stained 1% agarose gel after electrophoresis of λMSP-1 DNA digested with HindIII (lane 2), EcoRI (lane 3), BamHI (lane 4), HindIII plus EcoRI (lane 5), HindIII plus BamHI (lane 6), EcoRI plus BamHI (lane 7). The marker DNA (lane 1) is λc1857 DNA cut with HindIII. Sizes are 23.5, 9.5, 6.7, 4.4, 2.3 and 2.0 kb.

(b) Autoradiograph of the Southern blot of the above gel hybridised with the ^{32}P-labelled pMK-1 cDNA probe at 65°C in 3 x SSC. Restriction digests are as above. Hybridising bands in lane 6 show up upon longer exposure.

a



b

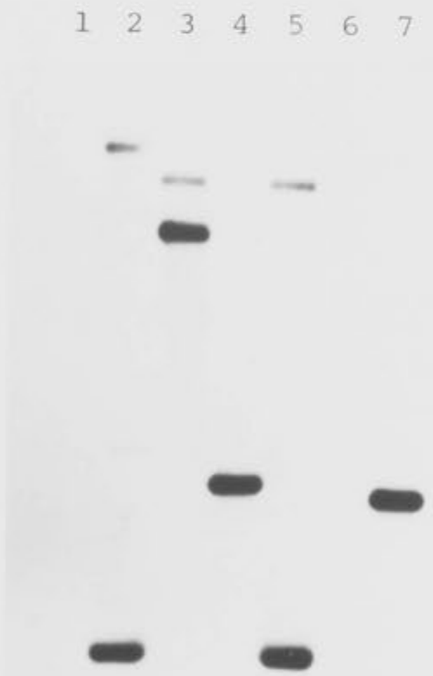
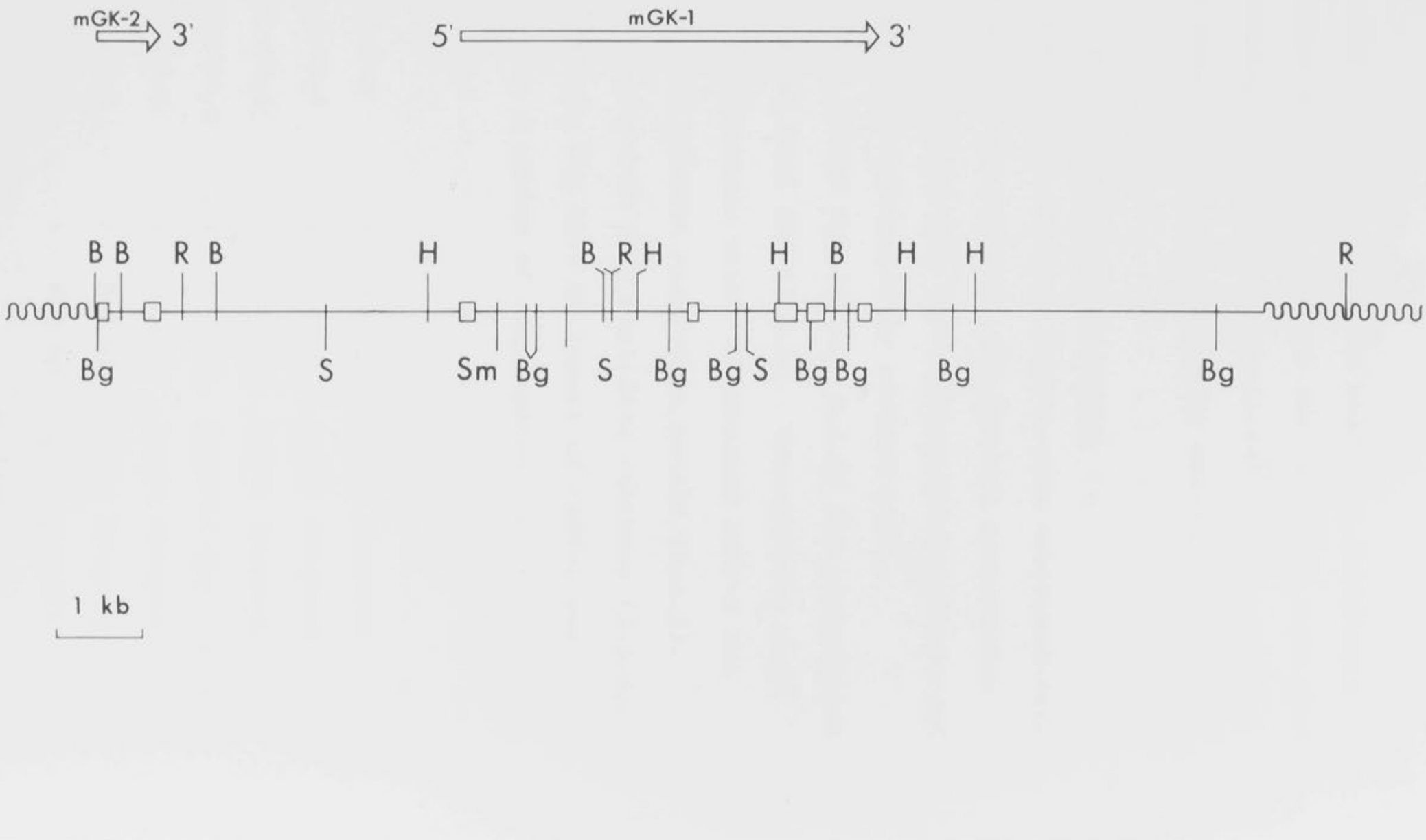


Figure 5.2 Restriction map of the bacteriophage lambda genomic clone, λMSP-1, which contains the kallikrein genes mGK-1 and mGK-2. Boxes denote exons of the respective genes mGK-1 and mGK-2. Both genes are transcribed in the direction indicated by the arrows. The position of restriction sites within the 13.5 kb insert are indicated by vertical lines. Abbreviations used are as follows: B-BamHI, R-EcoRI, S-SacI, H-HindIII, Sm-SmaI, Bg-BglII. The wavy lines represent the left and right arms of the lambda vector, Charon 28. The 5'BamHI site within the mGK-2 gene is the end point of the clone and therefore may only represent an MboI site in the genome.


mGK-2
3'
5'
mGK-1
3'
B B R B H B R H H B H H R
Bg S Sm Bg S Bg Bg S Bg Bg Bg Bg
1 kb

information for only approximately half of a kallikrein mRNA, the exons comprising each of the two genes were also mapped by northern blotting and confirmed by DNA sequence analysis as described in the following sections.

5.2.3 Subcloning of genomic sequences

Restriction fragments from λMSP-1 were subcloned into the plasmid pBR322. Generally, the desired restriction fragment was isolated from a low-melting point agarose gel (2.5.1), ligated to appropriately cleaved and dephosphorylated pBR322 DNA (2.3.7, 2.3.8) and transformed into the bacterial host RR1 (2.4.1). Recombinants were detected by hybridisation with ^{32}P-labelled λMSP-1 DNA fragments and by a plasmid miniscreen method (2.4.2). Plasmid DNA was prepared from amplified cultures (2.2.4).

In this manner the 13.5 kb insert of λMSP-1 was subcloned to give a series of overlapping clones, including:-

pK-2-B.3	:	300 bp	BamHI fragment
pK-2-B1.1	:	1.1 kb	BamHI fragment
pK-1-E5.0	:	5.0 kb	EcoRI fragment
pK-1-HSm.8	:	800 bp	HindIII-SmaI fragment
pK-1-B2.6	:	2.6 kb	BamHI fragment
pK-1-E8.0	:	8.0 kb	EcoRI fragment
pK-1-BH.8	:	800 bp	BamHI-HindIII fragment

The locations of the cloned DNA fragments in these recombinant plasmids is illustrated in Figure 5.5.

5.2.4 Exon mapping by northern blot analysis

As mentioned in 5.2.2, pMK-1 only cross-hybridises with exons 3, 4 and 5 of the gene mGK-1 and exons 4 and 5 of mGK-2. This is to be expected since pMK-1 only contains approximately half of the coding region for a kallikrein protease. The approach used to locate the remaining coding sequences of mGK-1 was to hybridise ^{32}P-labelled restriction fragments, from the gene region separating exons 3, 4 and 5 of mGK-1 and exons 4 and 5 of mGK-2 to northern blots of submaxillary gland mRNA.

Multiple samples of poly(A)$^+$ mRNA (1 µg) were fractionated on a 1% agarose gel and the RNA transferred to a nitrocellulose filter as described 2.6.2. The filter was cut into individual lanes and each was separately hybridised to the restriction fragments I-VIII as outlined in Figure 5.3. The restriction fragments which contained exon sequences hybridised to a common mRNA size class of approximately 950 nucleotides (Figure 5.4). This result demonstrates homogeneity in the mRNA size of the many different kallikrein genes expressed in this tissue. By this technique the 5' exon of mGK-1 was located to a 800 bp HindIII-SmaI fragment. All the exon-intron boundaries were precisely mapped by detailed sequence analysis as described below.

Figure 5.3 (a) Location of restriction fragments within the mGK-1, mGK-2 gene region, which were used to probe submaxillary RNA (see Figure 5.4). The DNA fragments designated I-VIII were prepared as hybridisation probes from the following plasmids:

I - pK-2-B.3

II - pK-2-B1.1

III - 2.6 kb BamHI-HindIII fragment isolated from pK-1-E5.0

IV - 0.8 kb HindIII - SmaI fragment isolated from pK-1-E5.0

V - 1.25 kb SmaI-BamHI fragment isolated from pK-1-E5.0

VI - 0.7 kb EcoRI-BglII fragment isolated from pK-1-B2.6

VII - 0.8 kb BglII doublet isolated from pK-1-B2.6

VIII - pK-1-BH.8.

(b) Physical map of 9.5 kb portion of λMSP-1 containing mGK-1 and mGK-2. Exons are denoted as boxes, and the direction of transcription of the genes is indicated by the heavy arrows. Abbreviations used are: B-BamHI, Bg-BglII, E-EcoRI, P-PstI, S-SacI, H-HindIII, Sm-SmaI, T-TaqI, Hp-HpaII.

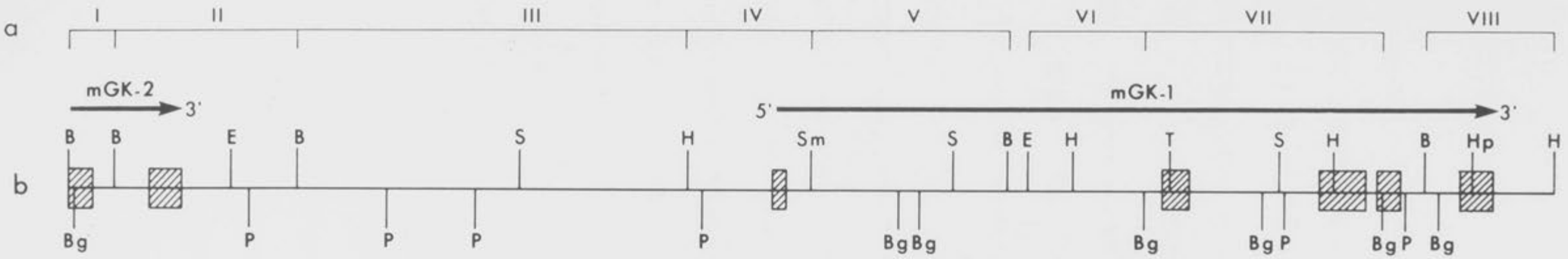
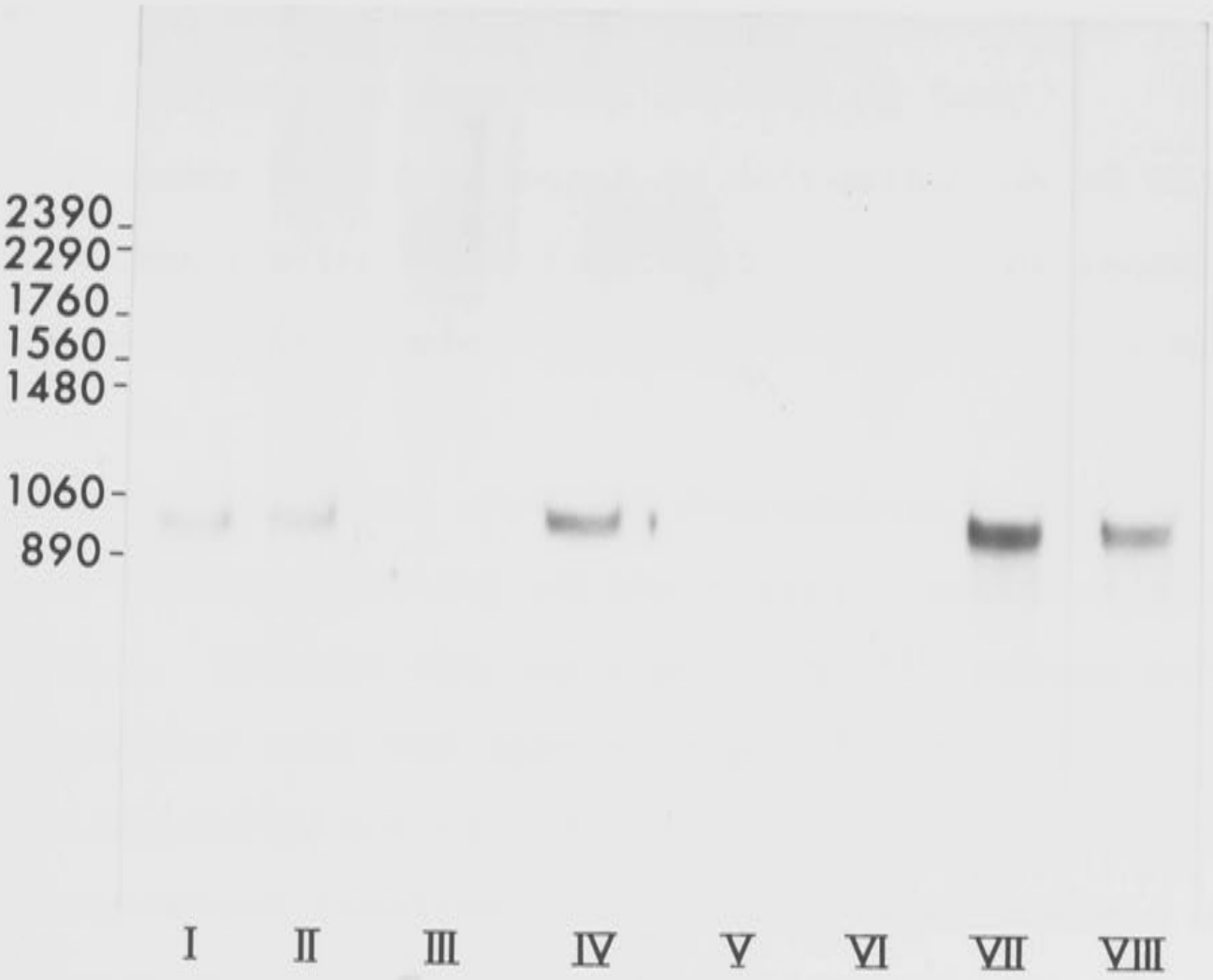

a
I
II
III
IV
V
VI
VII
VIII
mGK-2
3'
5'
mGK-1
3'
b
B
B
E
B
S
H
Sm
S
B
E
H
T
S
H
B
Hp
H
Bg
P
P
P
P
Bg
Bg
Bg
Bg
P
Bg
P
Bg

Figure 5.4 Northern blot analysis of poly(A)$^+$ mRNA isolated from male mouse submaxillary gland. Eight samples of poly(A)$^+$ mRNA (1 µg) were fractionated on a 1% agarose gel and transferred to a nitrocellulose filter as described in 2.6.2. Tracks I-VIII were individually hybridised at 42°C in 3 x SSC, 50% formamide for 36 hours with ^{32}P-labelled hybridisation probes prepared from the restriction fragments I-VIII, respectively (as illustrated in Figure 5.3). Filters were washed in 2 x SSC at 21°C and exposed for autoradiography at -70°C for 12 hours in the presence of an intensifying screen. The numbers alongside the autoradiograph represent the position of the eight RNA species from an influenza type A laboratory recombinant H3N2 strain (Sleigh et al., 1979) which were run as size markers. Restriction fragments I, II, IV, VII and VIII hybridised to a common mRNA size of approximately 950 bases. The lower intensity of hybridisation observed with fragments I and II were due to the lower specific activity of these two hybridisation probes.


2390
2290
1760
1560
1480
1060
890
I
II
III
IV
V
VI
VII
VIII

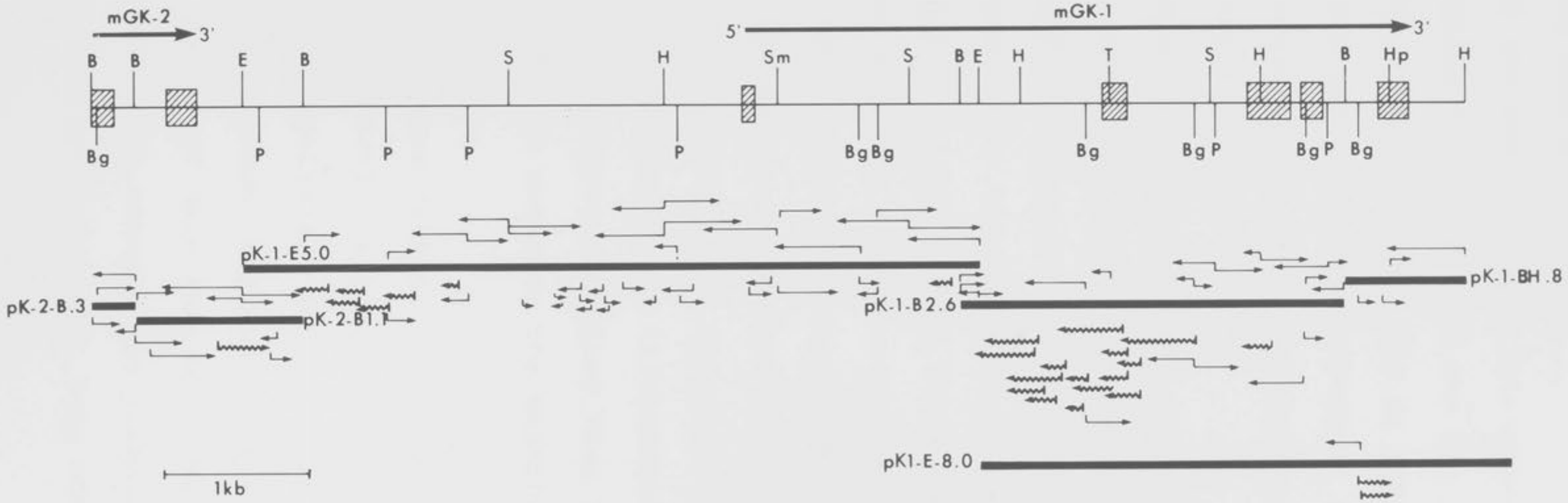
5.2.5 Sequence analysis of λMSP-1

The overlapping pBR322 subclones of λMSP-1 described in 5.2.3 and outlined in Figure 5.5 were used as a source of DNA for sequence analysis. Almost 9,500 contiguous base pairs of the mGK-1, mGK-2 gene region were determined by the strategy shown in Figure 5.5. The majority of the sequence was derived from both strands of DNA.

Sequences were determined by a combination of the chemical degradation method (2.7.1) and dideoxy sequencing methods using the single-stranded M13 phage vectors mp8 and mp9 (2.7.2).

Sequencing by the chemical degradation method was based on the availability of the restriction sites shown in Figure 5.5. Plasmid DNA was cleaved by one enzyme and 3'end labelled with the appropriate α-^{32}P-dNTP using reverse transcriptase (2.3.4a). PstI and SacI restriction sites were 3'end labelled using terminal transferase with α-^{32}P-cordycepin (2.3.4b). In each case the labelled fragments were then cleaved with another restriction enzyme(s) in order to produce fragments labelled at only one end and easily resolved by polyacrylamide or agarose gel electrophoresis. Purified labelled fragments were chemically modified and prepared for DNA sequencing gel electrophoresis as described in 2.7.1. The cleavage products of the DNA sequencing reactions were separated on thin 20%, 10% and 5% polyacrylamide denaturing (7.0 M urea) gels (2.5.4). Routinely up to 500 bases of sequence could

Figure 5.5 Sequencing strategy for the 9.5 kb mGK-1, mGK-2 gene region. The pBR322 subclones used for sequence determination are shown as solid lines below the restriction map of the 9.5 kb portion of λMSP-1. Sites on the restriction map are abbreviated as follows: B-BamHI, Bg-BglII, E-EcoRI, P-PstI, S-SacI, H-HindIII, Sm-SmaI, T-TaqI, Hp-HpaII. Only the TaqI and HpaII sites used for sequence analysis are shown. The arrows above and below the solid lines give the direction and length of sequence obtained. Sequences indicated by arrows above the subclones were obtained by the chemical degradation method. The arrows below the subclones indicate dideoxy sequencing of a defined (AluI, Sau3A, HaeIII, RsaI or BglII) restriction fragment subcloned into M13 mp8 or mp9. The arrows above the subclone pK-2-B1.1 in the area of overlap with pK-1-E5.0 indicate chemical degradation sequencing as do the arrows above pK-1-B2.6 in the area of overlap with pK-1-E5.0. Wavy lines below the subclones indicate sequence data obtained by cloning a series of Bal31 exonuclease-generated deletions into mp8. For more details on this method see 5.2.5.


mGK-2
3'
mGK-1
5'
3'
B B E B S H Sm S B E H T S H B Hp H
Bg P P P P Bg Bg Bg P Bg P Bg
pK-1-E5.0
pK-2-B.3
pK-2-B1.1
pK-1-B2.6
pK-1-BH.8
pK1-E-8.0
1kb

be resolved using this gel system.

Portions of the sequence were determined by cloning specific restriction fragments into the M13 vectors mp8 and mp9 (2.7.2). For example the two 800 bp BglII fragments from pK-1-B2.6 and the 600 bp PstI fragment from pK-1-E5.0 were partially sequenced by this technique as were other fragments outlined in Figure 5.5.

The DNA sequencing method termed "shotgunning" (Messing et al., 1981) was used to partially sequence the 1,050 bp SacI-HindIII fragment from pK-1-E5.0. Aliquots (2 µg) of this fragment were cleaved with the restriction enzymes AluI, HaeIII and RsaI. The blunt-ended cleavage products of these separate digests were ligated into HincII cleaved M13 mp8 DNA (2.3.8) and transformed into the host JM103 (2.4.1). The resultant 36 clones selected by hybridisation to the SacI-HindIII fragment were subjected to sequence analysis (2.7.2). As illustrated in Figure 5.5, the overlapping sequence obtained from 11 of these clones was sufficient to sequence the majority of this region on both strands.

Several regions were difficult to sequence due to the lack of suitable restriction sites for chemical degradation and dideoxy sequence analysis. To overcome this problem, a rapid method of 'non-random' sequencing based on the generation of a series of deletions with the exonuclease Bal31 was developed. For example the plasmid pK-1-B2.6, containing the 2.6 kb BamHI fragment from mGK-1, was

linearised within the pBR322 sequences by the endonuclease ClaI, digested with Bal31 and aliquots taken at 1 minute time intervals over a 4 minute period (2.3.2). By this procedure, it was possible to generate a range of defined deletions from the ClaI site towards the EcoRI site within the 2.6 kb BamHI insert. Under the conditions used the exonuclease removed 400 bp from each end of the linearised plasmid per minute, as monitored by agarose gel electrophoresis. These deleted plasmids were then cleaved with EcoRI in order to release fragments having one blunt-end due to Bal31 digestion and one cohesive end due to EcoRI cleavage. The fragments were then subcloned into EcoRI-HincIII-cleaved M13 mp8 DNA. Recombinant phage were detected by hybridisation to λMSP-1 DNA and sequenced using a synthetic DNA primer (2.7.2). As illustrated in Figure 5.5, the majority of the sequence of pK-1-B2.6 was determined by this technique.

The DNA cloned in the plasmids pK-2-B1.1, pK-1-E8.0 and pK-1-E5.0 were also partially sequenced by this method. The plasmid pK-2-B1.1 was linearised within the pBR322 sequences by the restriction enzyme SalI (the pBR322 SalI site is 170 bp distant from the BamHI site containing the 1.1 kb insert), and digested with Bal31 for 2, 3 and 4 minutes. Approximately 400 bp/minute was removed from each end of the plasmid, thus resulting in the loss of one of the BamHI sites. These shortened DNA fragments were then cloned into BamHI-HincII-cleaved mp8 DNA. Only one

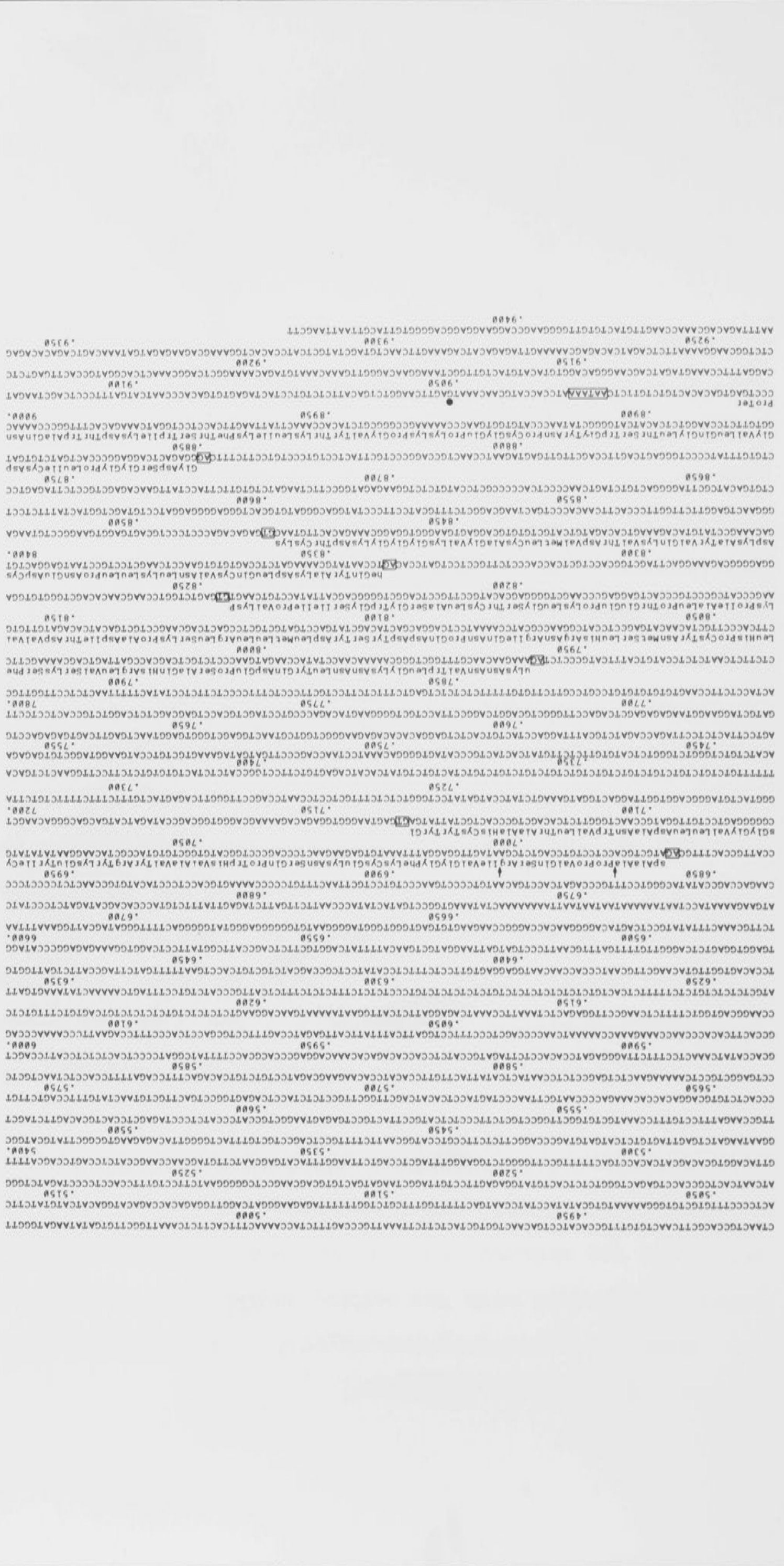
recombinant deletion clone was obtained as shown in Figure 5.5. This poor cloning efficiency could be explained by the incomplete digestion of the M13 vector with BamHI and HincII which was probably due to the closeness of these sites within mp8 (see Figure 2.2).

The DNA from the plasmid pK-1-E5.0 was linearised within the insert by the restriction enzyme SmaI. After Bal31 digestion for 1-6 minutes, the plasmids were cleaved with EcoRI. As this plasmid would still contain both EcoRI sites, a range of deletions extending towards both sites would be expected to result. The M13 recombinants obtained (see Figure 5.5) after ligation of these products into EcoRI-HincII-cleaved mp8 DNA, were screened by hybridisation with probes specific for the 5' and 3' ends of the pK-1-E5.0 insert (900 bp EcoRI - BamHI fragment from pK-2-B1.1 and the 100 bp BamHI-EcoRI fragment from pK-1-B2.6 respectively). In this way it was possible to determine the orientation of the M13 recombinants. Two deletion clones from the plasmid pK-1-E8.0 were constructed by initially cleaving the plasmid with BamHI, digesting with Bal31 for 2 and 3 minutes and releasing the required fragments by EcoRI digestion before ligating into EcoRI-HincII-cleaved mp8 DNA (Figure 5.5).

The complete 9,436 bp sequence of this region is presented in Figure 5.6 a and b.

Figure 5.6(a)(b) Sequence of the mouse mGK-1 and mGK-2 gene region. Only the sense strand is shown. The coding sequences which make up the five exons of the complete mGK-1 gene and 3' two exons of mGK-2 have been translated. The arrows in exon 2 of mGK-1 define the proposed boundaries between the signal peptide, zymogen peptide and the active protease. The GT-AG sequences at the intron-exon boundaries are boxed. The 'CAT' box (Benoist et al., 1980) and the variant 5'-TTTAAA-3' 'Goldberg-Hogness' box (Corden et al., 1980) preceeding mGK-1 are overlined. The two putative initiation sites of RNA transcription for mGK-1 are marked by triangles. The polyadenylation site sequences 5'-AATAAA-3' (Proudfoot and Brownlee, 1976) found in the 3'untranslated regions of mGK-1 and mGK-2 are boxed. Circles indicate the transcription termination sites. The TC-rich repeat in the intergenic space is underlined.

```
        GlnAsnAlaLysAspLeuGlnCysValAsnLeuLysLeuLeuProAsnGluAsnCysAspLysAsnHisAsnLysLysValThrAspValMetValCysValGlyGlu
GGATCCACAGTCCAAAATGCAAAAGATCTCCAGTGTGTGAACCTCAAGCTCCTGCCTAATGAGAATTGTGACAAAAACCACAATAAGAAGGTGACAGATGTCATGGTGTGTGTAGGAGAG
                                                 .50                                                  .100
MetAspGlyGlyLysAspThrCysVal
ATGGATGGAGGCAAAGACACTTGTGTGGTGAGACAGCCACTTTCTGCTGTGAGGTGAAGGGCTGAAAGCTGGAACTGAGGTTCTTGGTTCCCACTTCAACACCCTGACTAGGCAGGCTCT
                             .150                                              .200
TTTTATCCTTCCCTATGGAAGGGATGTGCACTGGGAGGGGAGGATCCTGTAGCTGGCATGATTTCTCACCTTAGACTTGTCTTAGGGGACAGTCTTGTGAGCCCCTCACCACAAACCAAT
         .250                                              .300                                                  .350
GTTCTCATACGATCTGAGGTCCCAAGAGCTAGATTCCCTCACCTACTGACTAGAGCTGCCTCTTAAGAGTCCCTGTGTTCCTCCCTGGGGTCAGTTCCAGGTTGTCAAGTAGGACTCCAA
                                       .400                                              .450
                                        GlyAspSerGlyGlyProLeuIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlySerIleProCysSerLysP
AGTCAGGTCCTCCATTACTGCTGTCCCTTTCCTTCTTTCAGGGAGACTCAGGAGGCCCACTGATATGTGATGGTGTTCTACAAGGTATCACATCATGGGGCTCTATCCCATGCAGTAAAC
                    .500                                              .550                                            600.
roAsnAlaProGlyIleTyrThrLysLeuIleLysPheAsnSerTrpIleLysAspThrMetThrLysAsnAlaTer
CCAATGCGCCGGGCATCTACACCAAACTTATTAAGTTTAACTCCTGGATAAAAGACACTATGACGAAAAATGCCTAAGTGTCACATTATCCTCTGTTCTGAATAAAATCCACTATGCAAC
    ●                                            .650                                              .700
AAATGAGTCAAGGTCTGACAGTCTCTATCCTACTGGAATACAGTACACAACCAAGGTCCCAACTAAGGTGAGGGCTTCCAAAGTAGATCAGCAAGGGACAGGTGTACATGTACTGTTGGT
                             .750                                              .800
TAAGGACATGGGGTTGAGACAAATGCAGACAAAAGGCTCAGACCAACTCAAGGATTCCACTTGAGTCTCCTCTGGCAAAGAAAATGTTCATATCACAGAGGAAAAGTCTAAAGAGATCAG
         .850                                              .900                                                  .950
AATGGGAGACTCCAGCTAGGCCTGATCCTATACTGGAAGCCGAAGGGATGATAAACAGTCAGACGCTCCGAATTCAGACTGCCTACCTGAGTTTCTTAGGTTGGGCTGGGGTGTTTTGTT
                                       .1000                                             .1050
AATACAGCCTTGGCTCGCTGGAGAAGAATTTGTGCACAGAAGGGCAGGGACTGCCAAAGGCTAGGTAGGCTGCAGTGTCTCTGCTCCCAGTGGCCACTGGAGGGGGTTTCTTTGGTAGGG
                    .1100                                             .1150                                          1200.
CTGGATCAAAGGATGAAGAGTGACTTAGAACCATGTCCTGTGGGCAGAGTAATACCCATCTTAGGAAAAGGGCCACACTGCTGCTGCACAGGACCACAAACCACTGTCTGTTTTAGAGGG
                                                 .1250                                             .1300
GACATAACCACAGCTAAGAGTATCAGAGATCCGAAGAAGGGCCAGAGAAGGTCAGAAAGATAGACAGAACTGGAAATCTACCAACAACTCTGGGTAGGAGGAGGCACTGGGGTTAAATGG
                             .1350                                             .1400
AGGATCCCCAGACACCCTTGTCCTTGCCAACACATATTGATTCCAATGCTAAGGAAACTGAAGATGAAGTCTATCTTCTTCAGGCAGCAGCAGACTGGTCGAAAGGAGAAGAACATGGTG
         .1450                                             .1500                                                 .1550
GGCAGGAGGATGAAGCATAGTTCCAGGGCCAGAATTATGAAGCCCTTGCTCATGTGAATGTGAACTGGGGAGGAAAGTTGGAAAGATAACTACAGTCATTAGTCATTCAAACCTCTCTCT
                                       .1600                                             .1650
CTCTCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCCCCCCTCTCTCTCTCCCTCCCCTCCCCAGTCTAGCTTCCTCTCTCCTTCCTCT
                    .1700                                             .1750                                          1800.
CACACCATTGGAGCACAACCTAAGTACTCCAGCACGCTGGAAGGGAGTGGGGGAATGTCTTTTCATGTTCTGACTGTGAAGTTAGAGACACATTGATCTTTTCTTGTTCTAAATACAAGG
                                                 .1850                                             .1900
TAAAAGCAGTCGAGGGTCTTGATGTATCTTTTTTTGGCAGAATATTTGCATACGTTGAGCAGCCTGCATTTCATTCTCAACACTGTGCTGAGCTGAGTCTGCAGCACATGCCTGTGATCC
                             .1950                                             .2000
CTGCAACTGGGAGGAGGATCAGAAAATCATGGTTGTCCACAGATACACATTGAGTTGGAGGTCAGCCTGGGCTAAACAGACATGTTACAAAGAGTGCAGAAGAAGGGAATGTAGCTGTCT
         .2050                                             .2100                                                 .2150
AGAATCCATCCCACAATGAGGGTCAGGAAATGGTTCTGAGTACAGCAAGTCCTAGAGTCTCAATAGGTGGAGTTCTTTGGGGGTGTGGGCTTCAAGTAGATAGAGACCCTGTGGAGTGTT
                                       .2200                                             .2250
TGTGTGGCGACTGGGTGCTGACTTCAGAAATTCCCATAAAAGAGGTATAAGAAAAACACCACCCTCCACTTCACATTAACAAAATCATTCCCTAGAAATTCTAAAGTAACTCTACTGCAA
                    .2300                                             .2350                                          2400.
CTCATTGAATGACCTACCATCGATGATGTCCCCACTGTGAGCCCAAACCTAGACCCCCAGAAAGAGTCCTGGGAGAGTGGGAGTATACCACCCACACAGCCACCTCTAGAATACAATGGA
                                                 .2450                                             .2500
ATGAACCCCCAGTCTATATATTCACTGACCAGAGATTTGTCCCACGTTGTCATATTTATGCCCCCTGCAGCCAATCCAGAGGCTATGAATCCAGTGTATGTTCGGTGCGCATAGGTATTC
                             .2550                                             .2600
ACACATCCATTATACTGAACACCAACAAGGTTACCGAGAGGATGAACATGAGGGACCAGTTTCGTAGGACAAGATTTCAGTCCTCCTCCCCAGGAATAGAGCAGGACAGCAGGGGAGGGA
         .2650                                             .2700                                                 .2750
AAAAATTCTGGGACCTGGATCTTTTCTCCCCTCTCTCTCAATGCATACAAACCTCCATCTTCTTCCTCATTTACATATTCCCCCTCATCTACAACCTGTTCCCTGTGGATGCTGTTTGGA
                                       .2800                                             .2850
GCTCGTCTCCAGCACCACTTCATCTCTTTCTATTCTCATGCACTGGACTCTTCCGGTGCAGTACACCTATCTCAAACTCCAGCATGATTATGTGTGCTTGAAGACATACAAAAAGATCAA
                    .2900                                             .2950                                          3000.
GTGCATGGATGTACTCACAAGTGGATGAATGCAGGCTTCCTCTCTTGTCCTTTAGAGCAAAGAACAGTGCAATTTTCAGAAAAACTTCAGCACAGAGGCCTAGGGCCAGAGGAGAACTTG
                                                 .3050                                             .3100
GTGGCCAGTTGATCACCAAGTCCTCACTTACCCTAGATTAGGGTTAGCCTAGCCTTCTTGGAGCTTACAGGGCACACAGGAATACACCCAAGTCAAATGTGCTAACTGTCATGAGAGCTG
                             .3150                                             .3200
AAATAGTCTAGATGCTCAACAAGAGGGGATGGGTCTGAAACCTGCAACAGTAAGTAGGTGTCAAAGGCTGGGTATTTCCCCTTCTCTGTACCAGAACCCATGCTGACTTCTTGATGTTTT
         .3250                                             .3300                                                 .3350
AACGTTTGAGACAAAGTCGTCTCCTGTGTAGTTTCACTGGCCTAAATACACTATGTGCACAGTCTAACCTGGACCTCACCAAGATCCACCAGCCCCTGACTCCTAAGTTCTGGGATCAAG
                                       .3400                                             .3450
GTGTGCACCAAAATATGCAGCTGCCAAGCTGAACCCTTAATGAGGCTCTTGGAGGGAAGTGACATTTTCTGGTTGCTGTTTCTTGAGAAAGGGTCTCACATGAACCTTGGCTTGGCTAAC
                    .3500                                             .3550                                          3600.
CTCAAAATGTAGCTCCTGATCTAGGAGATGAGGAAGGGAAAAGTCATAGTGACCCCAACATTGTTCAGGACAGGTGACACTCCCCTGGAATAACCTGAATTTTCATACTTTTGAATAAAA
                                                 .3650                                             .3700
GACTGATCAGGCCATGGTAATGTGCACCATAACTTTTATTCATTTTATTTTATAGGTTTGTTTTGAGACAGCTCTATATATAACCTTGGCTGTCTTGAAAGTCACTATGTAGACTAGGTT
                             .3750                                             .3800
GGTCTACATCACAGAATACTAGAGCCTAGAATCACACAGATCCACCTGGCTTCAAAGTACTGGTATTAGAGGAGTGCACTGCTATACACTGCTCCCAAGCTTCTCTTAGAAACAATAAAA
         .3850                                             .3900                                                 .3950
TCAGTGATCCAGAGTTCCAAGTCCAAAAAGGAAGCATTTTGGCATTCTCCAAGAAAGGATTTGTATCTGCAGGGAGCTTGCAGCCCTCTAGGCTATCACTTGTGAAAGACTCTGTTCTAA
                                       .4000                                             .4050
CTGCAACTTTCTGGGATAATCTTGGACTTGAGGTACAGGACCTTCCTTAAAGTTGTGGGTTAGGCAGCACTTGCTCCACTACCCACCATTCATACCAAGTCACCATTTCCACGACCTTGC
                    .4100                                             .4150                                          4200.
CATCTGGCTCCCAGTACCCGGAACACTAACAGTTCCAGAAACAGCAGGAATTGCCCTCCCAATCTCATTCTTGGGAATGTATCCAGGGTGACCTAGAGCCTTCTGGACAGTGGGTGCAAC
                                                 .4250                                             .4300
GAGGTCATCCCTTACCACAGCCAGGATGGTGCAAGTAGAGAATCCTAACCCTGTCAGCAAGGGCAGGGCTTTGTCAGCATCTGGGTGGCAACAGGACAGAGGAGGGGCTGTGGGGAGAAT
                             .4350                                             .4400
                              ▼         ▼                                   MetTrpPheLeuIleLeuPheLeuAlaLeuPheLeuGlyGlyIle
GGGGGATTTAAAGTCTCCCCAGGGAGCCTCAATAGCTCCAAGCTCACTGCCTGCTGCTCCTGAACACCTGTTACCATGTGGTTCCTGATCCTGTTCCTAGCCCTGTCCCTAGGAGGGATT
         .4450                                             .4500                                                 .4550
A
GGTGAGATGAGGGAATGGGAGGGGACACAGCCCTCACTCCCCTGCTTACTTAGGCCCTGGCTGTCCCTTTCCCCTTGCCCCTCTCCCCAGCTACATCATGAGCACCACCTGCTCCTCCTT
                                       .4600                                             .4650
TTCTGGCCCTGTACACTGCCAGGTCCCTCACATTCCTTGCTCCCGGGACAGGCCCCTGTTCCCTCCCACTTTCCAGTAATTTCTCCTCTTCCTTGGACACATAACCATCTCTGGTCTTCT
                    .4700                                             .4750                                          4800.
CCAGTCTTTCTAAGGAACCCAACCTCTTACCAAACTTAGGACCTCAGATTTCCCTACAGTTCATTATTTTTCCTAGAGCCTCATAACACAAACTCTAAGTCAGAGGCCTGCTTAAATCCT
                                                 .4850                                             .4900
```



5.2.6. Exon-intron assignment

When the previously-sequenced kallikrein cDNA clone pMK-1 was aligned with the coding sequences from mGK-1 or mGK-2, it was obvious that neither mGK-1 nor mGK-2 represent the genomic pMK-1 sequence. However, both are very closely related to pMK-1 and show 82-84% nucleotide sequence homology.

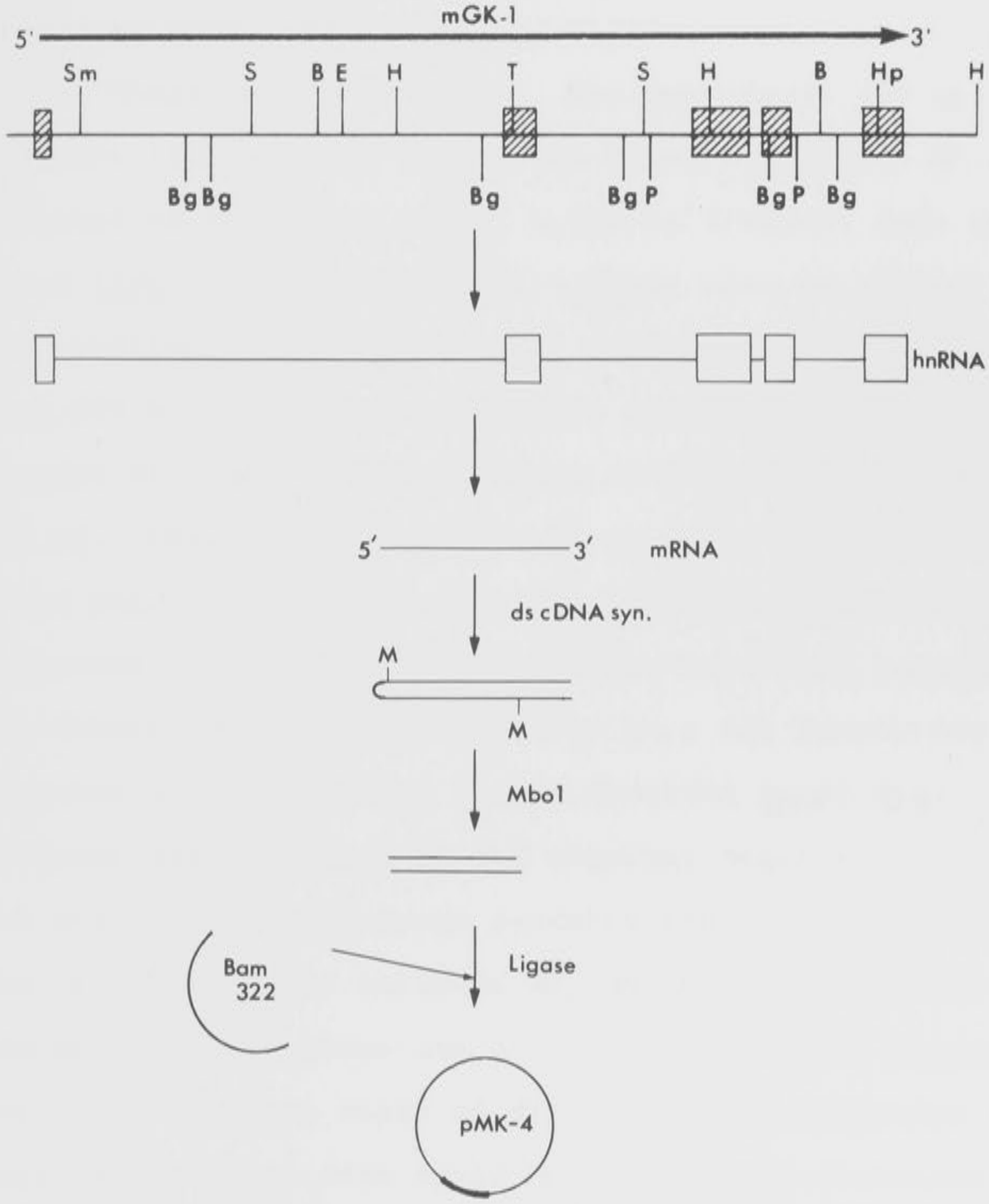
As previously mentioned, the sequences contained in pMK-1 cross-hybridise with exons 3, 4 and 5 of the genes mGK-1 and mGK-2. Comparison of these sequences facilitated the mapping of exon-intron boundaries. The exact location of exon 2 was assigned by a comparison of the amino acid sequence of γ-NGF with the predicted coding potential of mGK-1 in this region (see Figure 5.12 for comparison). All the intron sequences assigned (Figure 5.6 a and b) obey the 5'-GT-AG-3' consensus sequence (Lerner *et al.*, 1980). The sequence of γ-NGF is only for the active protease and no sequence data is available for the prepropeptides of kallikreins. The 5' exon of mGK-1, which contains a portion of the prepeptide, was located on an 800 bp *Hind*III-*Sma*I fragment, 2.4 kb 5' to exon 2 of mGK-1 by means of northern blot analysis (5.2.4). The boundaries of this exon and the 5' boundary of exon 2 were determined as outlined below.

5.2.7 Expression of mGK-1

In an attempt to both demonstrate the expression of mGK-1 and to map the 5'end of exon 2 and the 3' end of exon 1, a cDNA library was constructed from submaxillary gland poly(A)$^+$ mRNA. The cDNA library was constructed in a specific limited fashion in order to select for a mGK-1 cDNA clone spanning exons 1 and 4. The rationale behind this approach is illustrated in Figure 5.7.

Sites for the restriction enzyme MboI are present in exon I (nucleotide 4527) and exon 4 (nucleotide 8356) of mGK-1 (Figure 5.6 a and b). Double-stranded cDNA was synthesised as described in 2.3.6, cleaved with MboI and cloned into the BamHI site of pBR322. A library of 2,500 clones was constructed and subsequently screened by filter hybridisation (2.4.3) with an exon 1 (pK-1-HSm.8 insert) -specific probe from mGK-1. One of the positive clones, pMK-4, was selected for sequence analysis. The sequence of this clone was determined on both strands by chemical degradation and dideoxy methods (data not shown). The complete 500 bp sequence of pMK-4 agrees exactly with the exon sequences of mGK-1 between the two MboI sites (nucleotide 4527-8356). This result confirms the exon-intron assignments (Figure 5.6 a and b) for exons 1-4 and demonstrates that the mGK-1 gene is transcriptionally active in the submaxillary gland. At present it is not known whether the gene mGK-2 is also expressed.

Figure 5.7 Schematic representation of the strategy used for obtaining the cDNA clone pMK-4. Double-stranded cDNA was constructed using sequential reverse transcriptase reactions with submaxillary gland poly(A)$^+$ mRNA as template (2.3.6). To select for an mGK-1 cDNA clone, the ds cDNA was cleaved with the restriction endonuclease MboI. As MboI sites are present in exon 1 (nucleotide 4527) and exon 4 (nucleotide 8356) of mGK-1 (see Figure 5.6), this cleavage should result in the production of a 590 bp ds cDNA fragment corresponding to mGK-1. The MboI-cleaved ds cDNA (≈20ng) was ligated with 100ng of alkaline phosphatase-treated BamHI-cleaved pBR322 DNA (2.3.7, 2.3.8). The ligation mix was transformed into the E. coli host RR1, and the resultant 2500 clones screened by filter hybridisation (2.4.3) with an mGK-1 (exon 1)-specific probe. The sequence of one positive clone, pMK-4 corresponded exactly to the coding sequences of mGK-1 between the two MboI sites.


mGK-1
5'
3'
Sm
S
B E
H
T
S
H
B
Hp
H
Bg Bg
Bg
Bg P
Bg P
Bg
hnRNA
5'
3'
mRNA
ds cDNA syn.
M
M
Mbo1
Bam
322
Ligase
pMK-4

5.2.8 Transcription initiation site of mGK-1

Mung bean nuclease mapping experiments were carried out in an attempt to map the site of transcription initiation of mGK-1. To map the boundaries of exon 1, two fragments were labelled on the strand complementary to the mRNA. These were as follows: The PstI-Sau3A 499 bp fragment (nucleotides 4027-4527) from pK-1-HSm.8, 5' end labelled at the Sau3A site; a 224 bp fragment from pK-1-HSm.8 (Sau3A(nucleotide 4027) - EcoRI site in pBR322), 3' end labelled at the Sau3A site; a 5' end labelled EcoRI-TaqI 877 bp fragment (nucleotides 6104-7001) containing a portion of exon 2, isolated from pK-1-B2.6 (see Figure 5.6a and b). These templates were annealed to submaxillary gland poly(A)$^+$ mRNA and digested with mung bean nuclease as described in 2.8.1. The size of the protected fragments was determined by electrophoresis on a 10% denaturing sequence gel. A portion of the labelled EcoRI-TaqI fragment was sequenced by the chemical degradation method and the resultant cleavage products run as size markers. The results of this analysis are shown in Figure 5.8. It can be seen that there are no obvious protected fragments but rather a whole range of different size fragments in each lane. These data could be due to poor digestion by the mung bean nuclease of unprotected fragments, as well as a reflection of sequence heterogenity in both the coding regions and 5'untranslated region of different kallikrein mRNAs present in the submaxillary gland. Indeed, if each

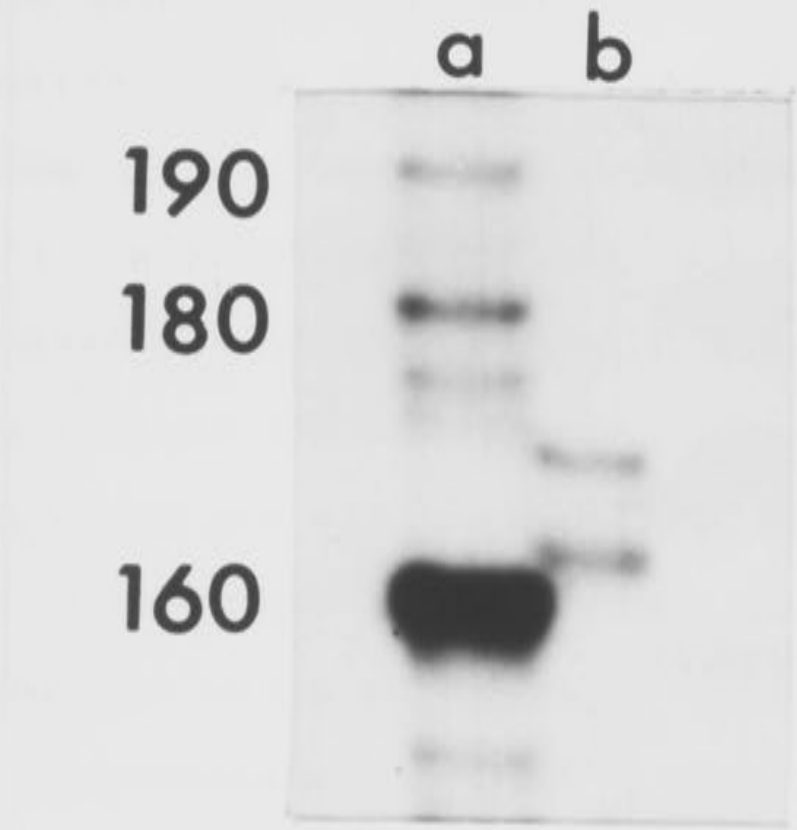
of the 25-30 kallikrein genes (see Chapter 6) is expressed in the submaxillary gland, such a result would be expected since the mung bean nuclease would cleave the template at any single base mis-matches.

To overcome this problem a primer extension experiment (2.8.2) using a short primer isolated from exon 2 of mGK-1 was performed. A 58 bp TaqI-RsaI fragment (nucleotides 7001-7059) was isolated from pK-1-B2.6, 5'end labelled and annealed with submaxillary gland poly(A)$^+$ mRNA as set out in 2.8.2. The primer was extended with reverse transcriptase and cold dNTP's and its size estimated by electrophoresis on a denaturing 10% polyacrylamide gel (2.8.2; Figure 5.9). When the length of the primer was aligned with the exon sequence of mGK-1, two putative capping sites 36 and 42 bp upstream from the initatior ATG were found (Figure 5.6 a). Given that there is a heterogeneous population of different kallikrein mRNA species in the submaxillary gland, each capable of annealing to the primer, it is impossible to determine if both mRNA sizes represent mGK-1 transcripts. The only conclusion that can be drawn is that there are two potential cap sites present in mGK-1. Interestingly, the sequence 5'-AGCTC-3' is present at both putative cap sites.

Figure 5.8 Mung bean nuclease mapping of the exon 1 and 2 boundaries of mGK-1. 877 bp EcoRI-TaqI fragment containing sequences from exon 2; a 499 bp PstI-Sau3A fragment containing sequences from exon 1; a 224 bp Sau3A-EcoRI fragment containing sequences from exon 1 were labelled on the strand complementary to the mRNA (2.8.1). These were annealed to submaxillary poly(A)$^+$ mRNA and digested with mung bean nuclease (2.8.1). Shown is an autoradiograph of a 10% sequencing gel used to size the products of digestion (Lane 1: EcoRI-TaqI fragment; Lane 3: PstI-Sau3A fragment; Lane 4: Sau3A-EcoRI fragment). A portion of the labelled EcoRI-TaqI fragment was sequenced by the chemical degradation method (2.7.1) and electrophoresed as size markers (lanes 2: left to right G,P,Y and C cleavage products).


3 4
2
1

Figure 5.9 Primer extension analysis. Autoradiograph of a 10% sequencing gel used to size extension products. The experiment was performed as outlined in 2.8.2 and 5.2.8. Lane b shows the position of the two extension products obtained and lane a shows the position of end-labelled HpaII fragments of pBR322. The size (in bases) of these fragments is indicated by the numbers alongside the autoradiograph.


a
b
190
180
160

5.2.9 Structure of a kallikrein gene

The mGK-1 gene spans 4,572 bp from the point of initiation of transcription (if 36 bp upstream from ATG) to the poly(A) addition site. There are four intervening sequences interrupting the coding region of the mGK-1 gene. All four introns begin with a GT dinucleotide and end with an AG, sequences thought to be necessary for correct RNA splicing (Lerner et al., 1980). The four introns are 2418, 821, 96, and 372 bp long, respectively, and the corresponding five exons code for 15, 54, 95, 46, and 51 amino acids. The mGK-2 gene exhibits a virtually identical 3'end structure to that of mGK-1. Intron D occurs in exactly the same position and separates exons 4 and 5 by a similar 374 bp. While these exons from the respective genes are 82% conserved in nucleotide sequence, the intron D sequences of both genes are 79% homologous when aligned for maximum homology.

The 3'untranslated regions of mGK-1 and mGK-2 (as predicted by analogy with the 3' untranslated region of pMK-1) are 45 and 48 bp respectively (Figure 5.6 a and b). The high degree of nucleotide sequence homology in exons 4 and 5 is retained in the 3'untranslated region. Allowing for a single 3 bp insertion in the mGK-2 untranslated region, the respective sequences can be aligned to give 91% sequence homology. Both genes possess the putative poly(A) addition signal sequence 5'-AATAAA-3' (Proudfoot and Brownlee, 1976) located 19 bp from the 3'

end of the mRNA (with respect to the pMK-1 mRNA sequence). The predicted size of the spliced transcript from mGK-1 is 873 nucleotides (if the cap site is 36 bp upstream from ATG). Northern blotting techniques show that kallikrein mRNA present in submaxillary gland is approximately 950 nucleotides in length (see Figure 5.4). The difference in size (873 compared with 950) can be accounted for by the presence of a poly(A) tail of between 50-100 bases, as observed for most eukaryotic mRNAs (Adams, 1977; Darnell, 1982).

5.2.10 Control sequences in the intergenic spacer region

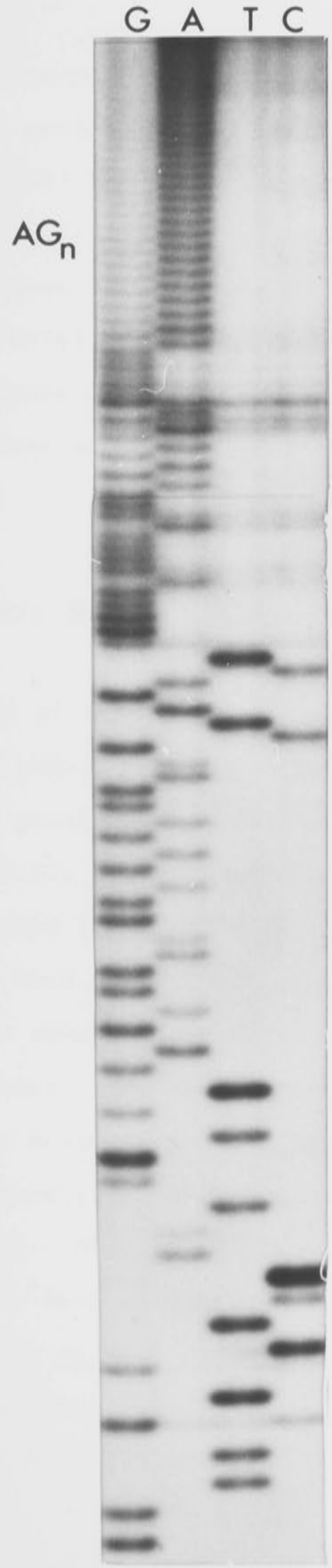
The complete nucleotide sequence of the 3,754 bp spacer region between mGK-1 and mGK-2 has been determined (Figure 5.6 a). This sequence should contain many of the sequences necessary for the expression of the mGK-1 gene. The two putative capping sites are preceded by a "Goldberg-Hogness" box (Corden et al., 1980) located 22 and 28 bp upstream. This sequence 5'-TTTAAA-3', also seen in the carboxypeptidase A gene (Quinto et al., 1982), is a variant of the consensus sequence 5'-TATAAA-3', which is thought to be important in determining the specificity of the initation of RNA synthesis by RNA polymerase II (Grosschedl and Birnstiel, 1980). In addition a 'CAT' box is located roughly 80 bp upstream from the cap sites (Benoist et al., 1980). This sequence is believed to be necessary for efficient RNA transcription (Grosveld et al., 1982).

The most striking feature of the spacer region is the presence of a TC-rich region 2,700 bp upstream from mGK-1 (Figure 5.6 a). A sequencing gel covering this region is shown in Figure 5.10. The simple T+C repeat is interspersed by occasional runs of C residues. This sequence is possibly repetitive DNA since the 580 bp BamHI-PstI fragment containing this sequence hybridises strongly when total mouse genomic DNA is used as a probe (data not shown).

5.2.11 Primary structure of the mGK-1 encoded kallikrein

The complete 261 amino acid sequence of the preprokallikrein encoded by mGK-1 is presented in Figure 5.11. The NH_2-terminal sequence of the primary translation product consists of a proposed hydrophobic signal peptide 18 amino acids in length. The assignment of the signal peptide was based on the observations that most signal peptides are between 17 and 30 amino acids in length and that the last residue before the cleavage point is often a small side chain amino acid (e.g: Ala, or Gly) (Chrétien and Seidah, 1981). Analogous to trypsin, the active kallikrein of 237 amino acids is preceded by a short 6 amino acid zymogen peptide (MacDonald et al., 1982). Kallikreins are known to be glycosylated; the typical glycosylation site of γ-NGF (Asn-Met-Ser) (Thomas et al., 1981) is also found in the same position (residue 102-104, Figure 5.11) in the mGK-1-encoded protease.

Figure 5.10 DNA sequence analysis of a portion of pK-1-E5.0. One of the Bal31 deletion M13 subclones derived from pK-1-E5.0 (see Fig. 5.5) was subjected to chain-termination sequence analysis. The products of this reaction were analysed by denaturing electrophoresis on a 10% sequencing gel. The sequence shown is of the complementary strand of nucleotides 1820≈ 1700 (Fig. 5.6a). The TC repeat referred to in 5.2.10 is thus represented on this gel as AG_n.


G A T C
AG_n

The coding potential of mGK-1 (Figure 5.11) predicts an active serine protease with a histidine (residue 41), serine (residue 189) and aspartic acid (residue 170) in the correct spatial configuration necessary for the formation of a serine protease catalytic site (Young et al., 1978; Krieger et al., 1974). The presence of aspartic acid at position 183 (Figure 5.11) also confers the specificity needed for cleavage at basic residues (Krieger et al., 1974).

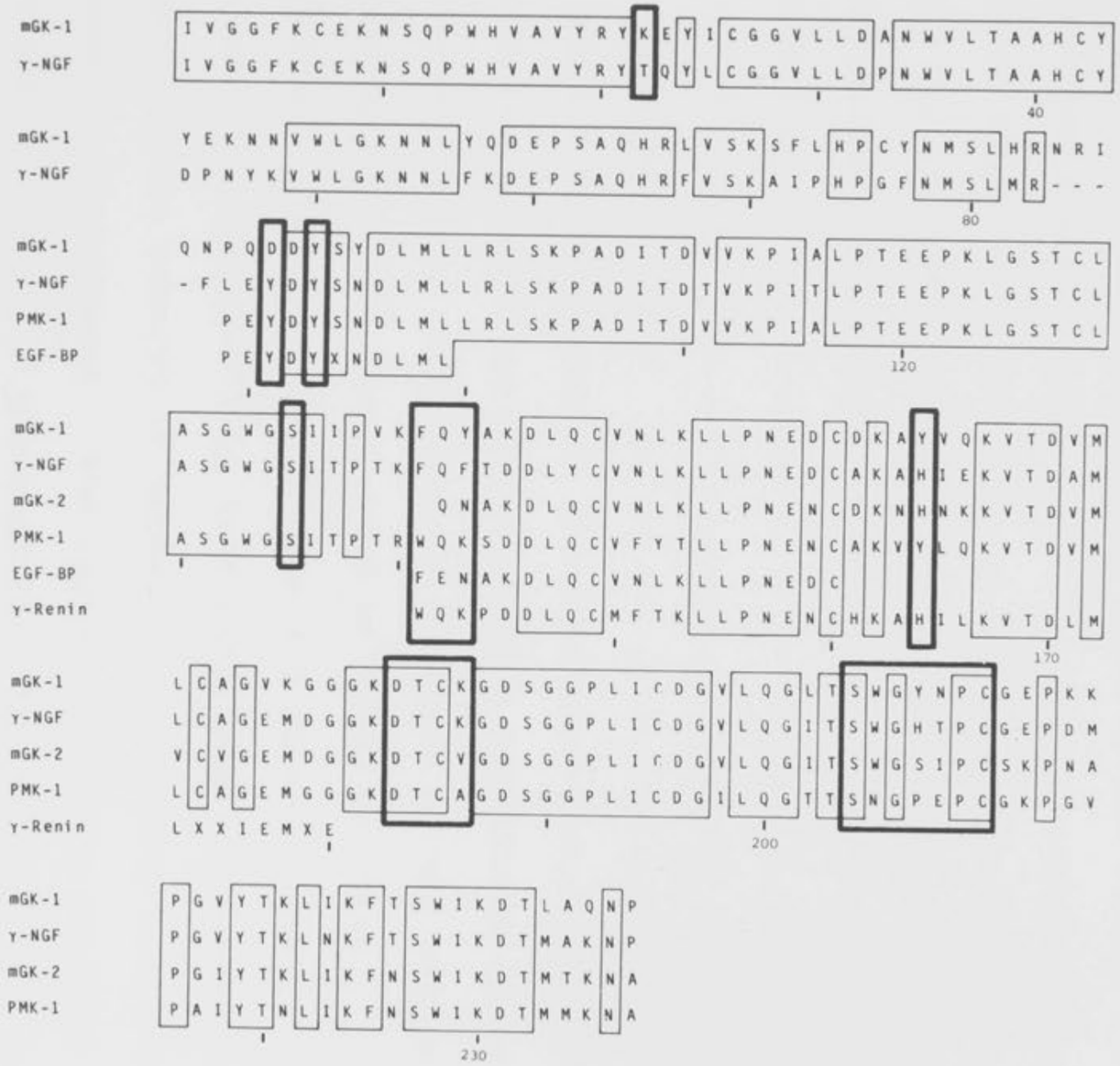
5.2.12 Comparison of mouse glandular kallikrein amino sequences

A comparison of the predicted amino acid sequences of mGK-1, mGK-2 and pMK-1 with the known amino acid sequence of γ-NGF and the partial sequences of EGF-BP and γ-renin is shown in Figure 5.12. As all of these proteases share extensive amino acid sequence homology of approximately 75%, it is clear that γ-NGF, EGF-BP and γ-renin are encoded by members of the same kallikrein multi-gene family as the clones isolated here. All these proteases are structurally very similar (for example, in the location of disulphide bridges) as previously noted for EGF-BP and γ-NGF (Server and Shooter, 1976). Even though EGF-BP and γ-NGF are highly homologous, they differ in their substrate specificities (Frey et al., 1979), a function attributed to alterations in the residues which make up the substrate binding pocket of serine proteases (Tschesche et al., 1979;

Figure 5.11 Exon sequences from the gene mGK-1. The coding region has been translated. The two putative initiation sites of RNA transcription for mGK-1 are at positions 1 and 7. The arrows indicate the proposed boundaries between the signal peptide, zymogen peptide and the active protease.

```
                                                                  MET TRP PHE LEU ILE LEU
AGC TCC AAG CTC ACT GCC TGC TGC TCC TGA ACA CCT GTT ACC ATG TGG TTC CTG ATC CTG
                                                       -24

                                                    ▼                    ▼
PHE LEU ALA LEU SER LEU GLY GLY ILE ASP ALA ALA PRO PRO VAL GLN SER ARG ILE VAL
TTC CTA GCC CTG TCC CTA GGA GGG ATT GAT GCT GCA CCT CCT GTC CAG TCT CGA ATA GTT
                                                                -1  +1

GLY GLY PHE LYS CYS GLU LYS ASN SER GLN PRO TRP HIS VAL ALA VAL TYR ARG TYR LYS
GGA GGA TTT AAA TGT GAG AAG AAC TCC CAG CCC TGG CAT GTG GCT GTG TAC CGC TAC AAG
                                                               20

GLU TYR ILE CYS GLY GLY VAL LEU LEU ASP ALA ASN TRP VAL LEU THR ALA ALA HIS CYS
GAA TAT ATA TGC GGG GGA GTC CTG TTG GAT GCC AAC TGG GTT CTC ACA GCT GCC CAC TGC
                                                               40

TYR TYR GLU LYS ASN ASN VAL TRP LEU GLY LYS ASN ASN LEU TYR GLN ASP GLU PRO SER
TAT TAT GAA AAG AAC AAC GTT TGG CTG GGC AAA AAC AAC CTA TAC CAA GAT GAA CCC TCT
                                                               60

ALA GLN HIS ARG LEU VAL SER LYS SER PHE LEU HIS PRO CYS TYR ASN MET SER LEU HIS
GCT CAG CAC CGA TTA GTC AGC AAA AGC TTC CTT CAC CCT TGC TAC AAC ATG AGC CTC CAT
                                                               80

ARG ASN ARG ILE GLN ASN PRO GLN ASP ASP TYR SER TYR ASP LEU MET LEU LEU ARG LEU
CGG AAC CGC ATC CAA AAT CCT CAG GAC GAC TAC AGC TAT GAC CTG ATG CTG CTC CGA CTC
                                                              100

SER LYS PRO ALA ASP ILE THR ASP VAL VAL LYS PRO ILE ALA LEU PRO THR GLU GLU PRO
AGC AAG CCT GCT GAC ATC ACA GAT GTT GTG AAG CCC ATC GCC CTG CCC ACT GAG GAG CCC
                                                              120

LYS LEU GLY SER THR CYS LEU ALA SER GLY TRP GLY SER ILE ILE PRO VAL LYS PHE GLN
AAG CTG GGG AGC ACA TGC CTT GCC TCA GGC TGG GGC AGC ATT ATA CCT GTC AAG TTC CAA
                                                              140

TYR ALA LYS ASP LEU GLN CYS VAL ASN LEU LYS LEU LEU PRO ASN GLU ASP CYS ASP LYS
TAT GCA AAA GAT CTC CAG TGT GTG AAC CTC AAG CTC CTG CCT AAT GAG GAC TGT GAC AAA
                                                              160

ALA TYR VAL GLN LYS VAL THR ASP VAL MET LEU CYS ALA GLY VAL LYS GLY GLY GLY LYS
GCC TAT GTA CAG AAA GTC ACA GAT GTC ATG CTG TGT GCA GGA GTG AAG GGT GGA GGC AAA
                                                              180

ASP THR CYS LYS GLY ASP SER GLY GLY PRO LEU ILE CYS ASP GLY VAL LEU GLN GLY LEU
GAC ACT TGT AAG GGA GAC TCA GGA GGC CCA CTG ATC TGT GAT GGT GTT CTC CAA GGT CTC
                                                              200

THR SER TRP GLY TYR ASN PRO CYS GLY GLU PRO LYS LYS PRO GLY VAL TYR THR LYS LEU
ACA TCA TGG GGC TAT AAC CCA TGT GGT GAA CCC AAA AAG CCG GGC GTC TAC ACC AAA CTT
                                                              220

ILE LYS PHE THR SER TRP ILE LYS ASP THR LEU ALA GLN ASN PRO ***
ATT AAG TTC ACC TCC TGG ATC AAA GAC ACT TTG GCC CAA AAC CCC TGA GTG ACA CAC TGT
                                                          237

CTG TTC TCA ATA AAA TCC ACC ATG CAA CAA ATG
```

Figure 5.12 Comparison of the coding potential of mGK-1 and mGK-2 with other mouse kallikrein amino acid sequences. The predicted amino acid sequence of mGK-1 and mGK-2 is aligned with the amino acid sequence of γ-NGF (Thomas et al., 1981) and the partial sequences of EGF-BP (Silverman, 1977), γ-renin (Poe et al., 1983) and the predicted sequence from the cDNA clone pMK-1 (Chapter 3). The boxed areas denote regions of the sequences which have identical residues. From the alignment of the sequences it seems that four amino acids have been removed from γ-NGF during maturation (see Discussion 5.3). The residues believed to line the substrate binding pocket are heavily boxed (Tschesche et al., 1979; Stroud et al., 1975; Hartley and Shotton 1971). Residues are numbered from the beginning of the active protease.


mGK-1
γ-NGF
mGK-1
γ-NGF
mGK-1
γ-NGF
PMK-1
EGF-BP
mGK-1
γ-NGF
mGK-2
PMK-1
EGF-BP
γ-Renin
mGK-1
γ-NGF
mGK-2
PMK-1
γ-Renin
mGK-1
γ-NGF
mGK-2
PMK-1
40
80
120
170
200
230

Stroud et al., 1975; Hartley and Shotton, 1971). By analogy with the overall structure of trypsin and elastase, the residues that line the substrate binding pocket can be predicted (Tschesche et al., 1979; Stroud et al., 1975; Hartley and Shotton, 1971). The kallikreins γ-NGF, EGF-BP and γ-renin, as well as those encoded by mGK-1, mGK-2 and pMK-1, show a significant variation in these particular residues compared to the rest of the sequence (Figure 5.12). It is therefore tempting to speculate that each of these kallikreins exhibits a different substrate specificity.

5.3 Discussion

This chapter describes the isolation and characterisation of two closely linked kallikrein genes (mGK-1 and mGK-2). The complete sequence of the region containing mGK-1 and the 3' half of mGK-2 has been determined. The gene mGK-1 spans 4.5 kb and is split into five exons totalling 873 bp in length, and is therefore typical of most eukaryotic genes, in possessing a mosaic exon-intron arrangement. The region 5' to the cap site of mGK-1 exhibits some of the signals ('TATAA' and 'CAT' box) known to be necessary for eukaryotic gene expression (Grosschedl and Birnstiel, 1980; Benoist et al., 1980). Although no known protein homologous to the mGK-1-encoded protease has yet been isolated, the finding of a cDNA clone identical to the coding sequence of mGK-1 indicates that

this gene is expressed in the male mouse submaxillary gland. The very similar chemical and physical properties of the mouse submaxillary gland kallikreins has made their characterisation at the protein level difficult and probably explains why no protease completely homologous to mGK-1 has as yet been characterised. The use of recombinant DNA technology has thus allowed many of the usual problems encountered in analysis of highly homologous related peptides to be circumvented.

The 3' end of another kallikrein gene (mGK-2) is found 3.7 kb upstream from mGK-1. This demonstrates that at least some of the kallikrein genes are tightly linked in the genome. The close linkage of kallikrein genes is further demonstrated by the finding of other genomic clones containing more than one kallikrein gene (Chapter 6). The 3' end of mGK-2 exhibits an identical exon-intron arrangement to that of mGK-1. Exons 4 and 5 are 82% homologous to their respective counterparts in mGK-1. The intron separating them is also highly conserved in sequence (79%). All the kallikrein genes which have so far been examined in detail (Chapter 6) appear to exhibit an identical exon-intron arrangement.

The mGK-1-encoded protease is initially synthesised as a preprokallikrein of 261 amino acids. The presence of a signal peptide has not been previously demonstrated for the kallikreins but is expected on the basis of their localisation in secretory granules (Pasquini et al., 1974).

Analogous to other serine proteases (Neurath and Walsh, 1976), kallikreins possess a short 6 amino acid zymogen or activation peptide. This prediction is consistent with the isolation of an inactive kallikrein which can be proteolytically activated without significant change in molecular weight (Schachter, 1980). γ-NGF isolated from mouse submaxillary gland is either a two chain or three chain species (Burton and Shooter, 1981) whilst EGF-BP is exclusively a three chain protein (Server and Shooter, 1976). Chain breakage is believed to be an auto-catalytic event with the subsequent removal of some residues by a carboxypeptidase B-like enzyme (Burton and Shooter, 1981). The alignment of the sequences of mGK-1 and γ-NGF shows an additional 4 amino acid residues (83-86, Figure 5.12) in the kallikrein autolysis loop where chain breakage in γ-NGF is known to occur (Thomas et al., 1981). This result suggests that there are four amino acid residues removed from the γ-NGF molecule during chain breakage. Kallikrein-like enzymes possess an extended autolysis loop of 11 additional amino acids (82-92, Figure 5.12) relative to trypsin. This kallikrein autolysis loop is on the surface of the enzyme and would be expected to extend far into solution, readily explaining the observed sensitivity of γ-NGF and EGF-BP to proteolytic cleavage (Bode et al., 1983).

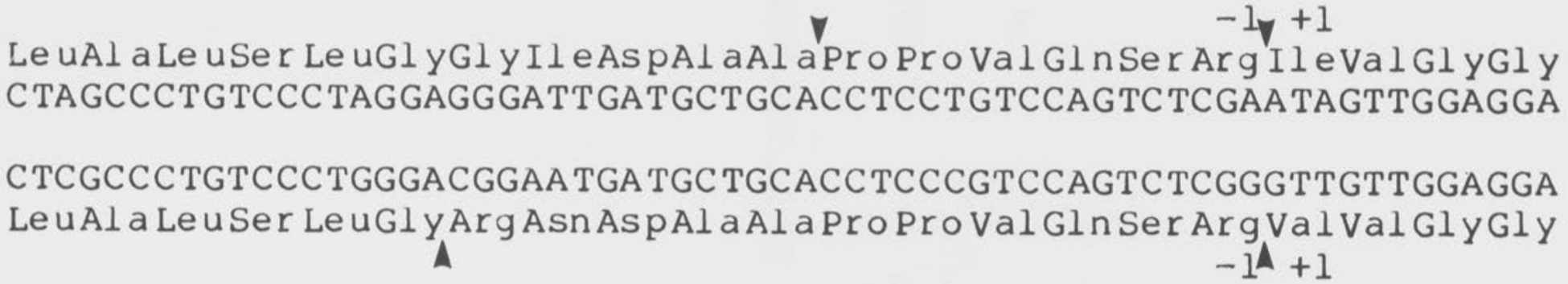
Recently, Swift et al., (1982) described the isolation and sequence analysis a cDNA clone encoding a rat

pancreatic kallikrein. The nucleotide sequence of the mGK-1 gene is 77% homologous with the corresponding sequence of the rat cDNA. The only difference in the overall structure of these coding sequences is at their 5' ends, where the rat cDNA contains an additional AUG. As illustrated in Figure 5.13, the rat mRNA appears to code for an additional four amino acids at the NH_2-terminus, owing to a postulated translation initiation at the first AUG. Consequently, Swift et al. (1982) have assigned the cleavage point between the signal peptide and zymogen peptide to a different position from that predicted for the mGK-1 protease. Owing to the overall homology between both these sequences, it is highly unlikely that the rat kallikrein would be so markedly different from the mouse kallikrein gene in this position; translation is therefore likely to initiate at the second AUG in the rat mRNA. Analysis of the amino-terminal end of rat preprokallikrein would resolve this question.

Each of the mouse kallikrein-type sequences (Figure 5.12) so far examined differs at amino acid residues implicated in the substrate specificity of serine proteases. This may reflect functional differences between most, if not all, of the 25-30 members of the kallikrein gene family (Chapter 6).

Figure 5.13 Comparison the 5'end structure of mGK-1 and a rat pancreatic kallikrein cDNA clone (Swift et al., 1982). The mouse sequence (top line) and rat sequence have been translated. As shown, Swift et al. (1982) assigned the translation initiation point at the first ATG in the rat cDNA sequence. The proposed boundaries between the signal peptide, zymogen peptide and active protease for the mGK-1-encoded protease and rat pancreatic kallikrein are indicated by arrows.

```
                                                        -24
                                                        MetTrpPheLeuIleLeuPhe
AGCTCCAAGCTCACTGCCTGCTGCTCCTGAACACCTGTTACCATGTGGTTCCTGATCCTGTTC

      AAGCTCAGCACCTGCTGCTCCTGCATGCCTGTTACCATGTGGTTCCTGATCCTGTTC
                                        MetProValThrMetTrpPheLeuIleLeuPhe
                                        -28
```



CHAPTER SIX

THE GLANDULAR KALLIKREIN GENE LOCUS

6.1 Introduction

The demonstration that all of the kallikrein genes in the mouse are located on chromosome seven and that two members of this gene family are present in a single lambda clone spanning only 13.5 kb suggest that all the individual members of the kallikrein gene family may be tightly linked in one locus on chromosome seven. In an attempt to determine if other members of the family were also as closely linked as mGK-1 and mGK-2, further bacteriophage lambda genomic isolates were analysed in detail. Furthermore, detailed analysis of the structure of the mGK-1 gene (Chapter 5) facilitates the selection of exon-specific hybridisation probes which could be used to ascertain whether other kallikrein genes possess the same exon-intron arrangement as that of mGK-1. The results presented in this chapter show that close linkage is a common feature of members of this gene family. In addition, members so far examined all appear to possess similar exon-intron arrangements.

Since hybridisation analysis of restriction digests of genomic DNA (Figure 4.1) has demonstrated that many, if not all, members of this family are highly homologous at the DNA level, it is impossible to distinguish between individual kallikrein genes by using extensive hybridisation probes such as pMK-1. For this reason the use of a chemically synthesised oligonucleotide probe homologous to only a portion of mGK-1 was investigated. This study

demonstrates the usefulness of synthetic oligonucleotides in the investigation of the expression of individual members of the kallikrein gene family.

6.2 Results

6.2.1 Characterisation of λMSP-2

Another kallikrein genomic clone (λMSP-2) was isolated from a genomic library derived from Quackenbush mouse as described in 2.2.7 and cleaved with different combinations of the restriction endonucleases BamHI, EcoRI, HindIII, SacI and SmaI. Following agarose gel electrophoresis the DNA was transferred to a nitrocellulose filter and hybridised either to a ^{32}P-labelled pMK-1 probe (specific for exons 3, 4 and 5) or to a probe prepared from the subclone pK-1-HSm.8 (specific for exon 1). An autoradiograph obtained after hybridisation with a pMK-1 probe is shown in Figure 6.1. The exon 1-specific probe did not hybridise with λMSP-2 DNA.

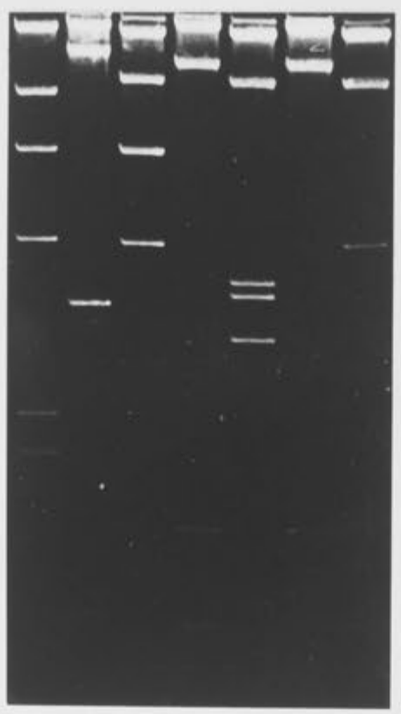
The restriction map of λMSP-2 (Figure 6.2a) shows the presence of a kallikrein gene (mGK-6) organised in a fashion analogous to that of mGK-1. The structure of mGK-6 was determined by hybridisation analysis with exon-specific probes (see above) and confirmed by sequence analysis as detailed in the following section. The clone only contains exons 2-5 of a kallikrein gene. The failure of the exon

Figure 6.1 Restriction and Southern blot analysis of λMSP-2.

(a) Ethidium bromide stained 1% agarose gel after electrophoresis of λMSP-2 DNA digested with HindIII (lane 2), EcoRI (lane 3), BamHI (lane 4), HindIII plus EcoRI (lane 5), HindIII plus BamHI (lane 6), EcoRI plus BamHI (lane 7). The marker DNA (lane 1) is λc1857 DNA cut with HindIII. Sizes are 23.5, 9.5, 6.7, 4.4 2.3 and 2.0 kb.

(b) Autoradiograph of the Southern blot of the above gel hybridised with the ^{32}P-labelled pMK-1 cDNA probe at 65°C in 3 x SSC.

a



b

1-specific probe to hybridise with λMSP-2 DNA could be due to a lack of homology between the 5' exons of the genes mGK-1 and mGK-6. This seems unlikely since the exon 1-specific probe cross-hybridises strongly with kallikrein mRNA (Figure 5.4), and with all other kallikrein genes so far analysed (see later). If exon 1 of mGK-6 was in a similar position to that observed for mGK-1 (i.e. 2400 bp 5' to exon 2) one would not expect λMSP-2 to contain the 5' end of mGK-6. As shown in Figure 6.2a exon 2 is located only 2 kb from the 5' end of λMSP-2. The isolation of another λ clone containing mGK-6 has shown exon 1 of this gene to be 2.4 kb distal to exon 2 (6.2.4).

6.2.2 Sequence analysis of mGK-6

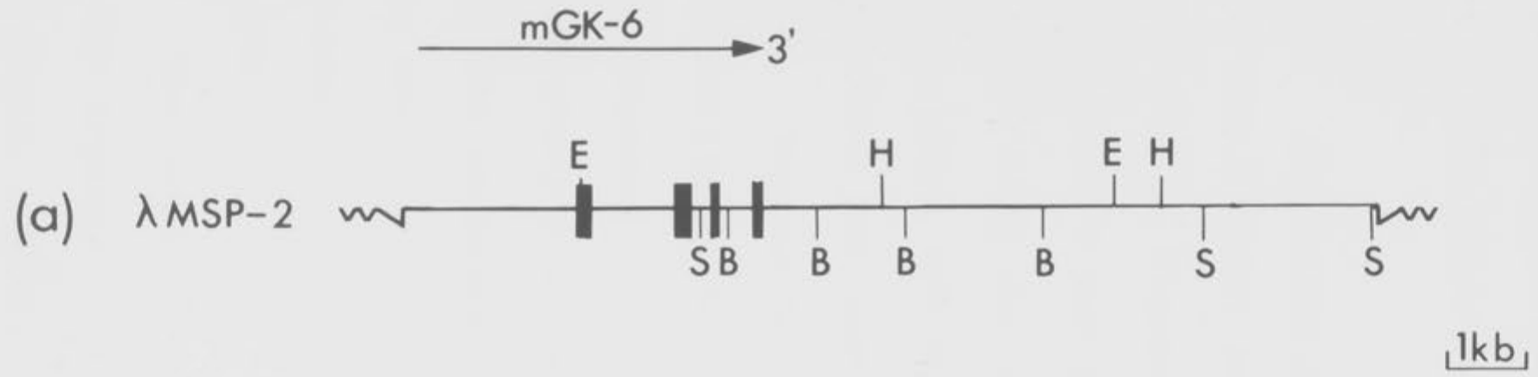
To obtain limited DNA sequence of mGK-6, especially around the exon-intron junctions, the 3.7 kb EcoRI-HindIII fragment (Figure 6.2), containing exons 2-5 of mGK-6, was subcloned into pBR322. DNA prepared from this plasmid, pK-6-EH3.7, was used for DNA sequence analysis.

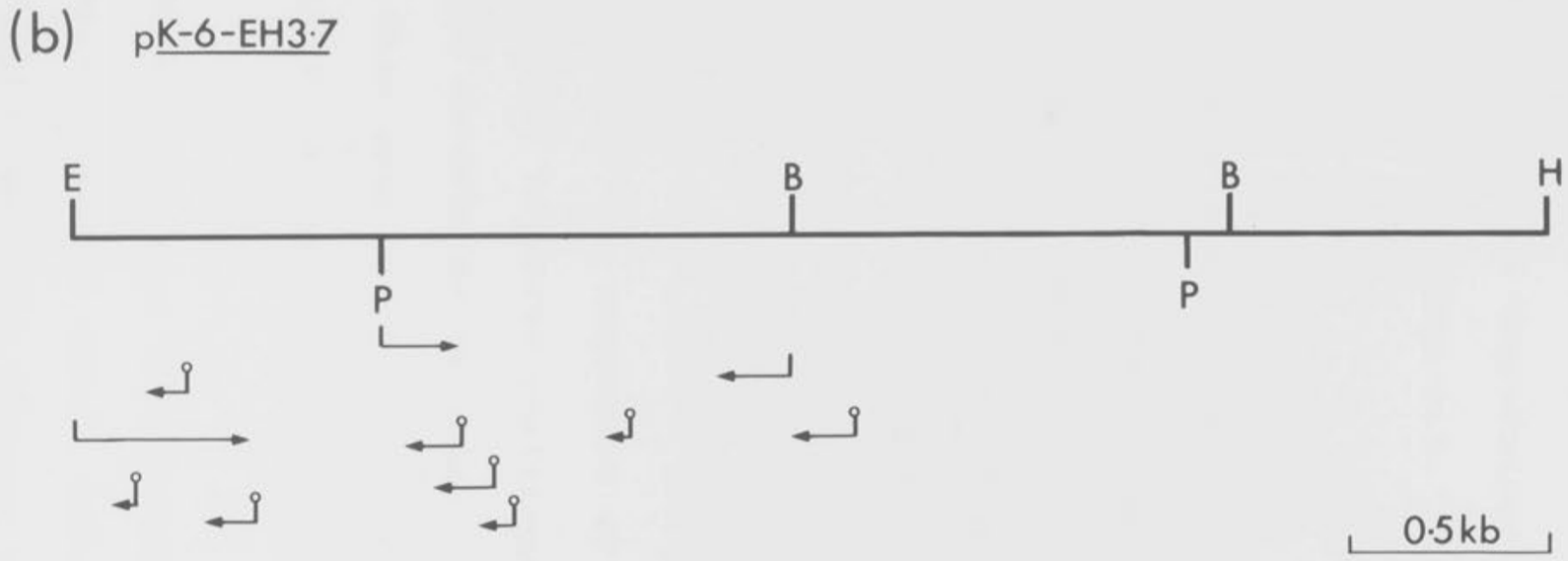
Sequencing data was obtained by the chemical degradation method and also by using the Bal31 exonuclease procedure (Figure 6.2b). The plasmid pK-3-EH3.7 was linearised by the endonuclease HindIII, digested with Bal31 and aliquots taken at 1 minute time intervals over a 6 minute period (2.3.2). The deleted plasmids were then cleaved with EcoRI and the required fragments subcloned into EcoRI-HincII-cleaved M13 mp8 DNA. Clones selected on

Figure 6.2

(a) Restriction map of λMSP-2. The 12 kb insert was mapped for the presence of kallikrein sequences as described 6.2.1. The boxes denote exons 2-5 of the kallikrein gene mGK-6 present on λMSP-2. The arrow indicates the direction of transcription. Wavy lines indicate the left and right arms of λ vector Charon 28.

(b) The 3.7 kb EcoRI-HindIII fragment containing the coding region from mGK-6 was subcloned into pBR322. The restriction map of this subclone pK-6-EH3.7 is shown. Restriction sites are: E-EcoRI, S-SacI, B-BamHI, H-HindIII, P-PstI. The arrows below the subclone indicate the direction and length of sequence obtained by the chemical degradation method. Circled arrows indicate sequence obtained by cloning a series of Bal31 exonuclease-generated deletions into mp8.


mGK-6
3'
(a) λMSP-2
E
H
E H
S B
B
B
B
S
S
1kb


(b) pK-6-EH3·7
E
B
B
H
P
P
0·5kb

the basis of specific hybridisation to λMSP-2 DNA were analysed by dideoxy sequencing (2.7.2).

The sequence obtained from the analysis of eight Bal31 deletion clones and from several chemical degradation sequencing experiments is presented in Figure 6.3.

6.2.3 Structure of the mGK-6 gene

The partial sequence of the subclone pK-3-EH3.7 presented in Figure 6.3 includes that of exons 2,3 and 4 and introns B and D of the mGK-6 gene. Comparison of this sequence with the previously described sequences of mGK-1, mGK-2, pMK-1 and the amino acid sequence of γ-NGF suggest that mGK-6 may represent the gene for the cDNA clone pMK-1. The nucleotide sequences corresponding to amino acid residues 150-158 and 206-210 of mGK-6, pMK-1 and mGK-1 are compared in Figure 6.4. Although the sequences of mGK-6 and pMK-1 are identical and differ at three positions from the sequence of mGK-1, more nucleotide sequence data from mGK-6 is required before this gene can be definitively assigned as the genomic equivalent of the sequences present in pMK-1.

The sequence data from mGK-6 shows that the position of the introns is exactly conserved between mGK-6, mGK-1, and mGK-2. In mGK-6, an intervening sequence (intron B) of approximately 1150 bp splits the codon for amino acid 69, as does the 841 bp intron B of mGK-1. Although the precise boundaries of intron C have not been located by sequence

Figure 6.3 Partial DNA sequence of mGK-6. The regions that have been translated are from exon 2 (residues 34-69), exon 3 (residues 70-85 and 150-158) and exon 4 (residues 206-210). Numbers refer to the position in a 261 amino acid preprokallikrein. Exon-intron boundaries are boxed.

```
                                  40                                  50
 AsnSerGlnProTrpGlnValAlaValTyrArgPheThrLysTyrGlnCysGlyGlyIleLeu
GAATTCCCAGCCCTGGCAAGTGGCTGTGTACCGCTTCACCAAATATCAATGTGGGGGTATCCTG
                            60
LeuAsnValAsnTrpValLeuThrAlaAlaHisCysHisAsnAs
CTGAACGTCAACTGGGTTCTCACAGCTGCCCACTGCCATAATGAGTGAGTAAGGGTGGAAACAG

GAAAGCAGGGTGGCAGCCAGAGAACATGACTCCAAGACAGGCTGGGGTACTGAAGGGCAGAGTG

GGGGACTGGCTGAAAGTCTATCAATGGCCTCCTGGTTCTCTATTTGCTCCTCCATCCAGGTTTA

TGTACAGAGTCCTGTTACTTTCTTTTCTGTATTATTGTCTGTGTATCTCCTGTCTGTATTCACA

TTTGAGTGTCTCTGTGGTCATCTGTCACTGTGTGTCTCTGCCTTGGGACACTGGGCACACCTTT

CTGGTTCTGGGTCTCATGTGTTTTTTTATCACTACTGCCCTAGTGGGGACCAGGACTTTAGGGC

CCC----350bp----CTGCAGGCCAGGCCTTACCTGCTGGGGAAGTTCAGCCCTGTCCTAGCT

GCACCTGAGCAGCTCTCAGGCCTGCCCCTCCTCCCTGCCCCCTTTTACTTTGTCTTCTGTTCCT

GTCTCTGTCTCTTTCAGCCTCTCCGACTCTATATGTATTTGTGTCTGTCTGTCTGTCTGTCTTT

CCGTGTTTCTGTCTCTGAGTCTCTCTTCTACCTCCCTCCCTTCCCCGACTCTTCCCCTTACTCT
                                                                  70
                                                              pLysTyr
GCCCAATCTCCCTGGTTGCCTCCTCTCACTCCCTTCATCTCCTTCCCTGACCCCCAGCAAGTAC
                                 80                              150
GlnValTrpLeuGlyLysAsnAsnPheLeuGluAspGluPro                  SerThrCys
CAGGTGTGGCTGGGCAAAAACAACTTTTTGGAGGATGAACCC----195bp----AGCACATGC
                                                       210
LeuAlaSerGlyTrpGly                    LysAspThrCysAla
CTAGCCTCAGGCTGGGGCA----240bp----AAAGACACTTGTGCGGTAAGACAAACCCCCTC

TGCAATGAGGTGAAGGGCTGAGAGAACTGAGGTTCTTGGTTCCCAGTTGAACACCCTGGCTAGG

CAGGCTCTTTGCACCTTCCCTATGGAAGGGA----17bp-----GGATCCTGTACCTGGTACTA

TTTCTCGCTCTGTGACTTCGCTTAGGGGAATGTCGACCGATCCACCACTCACCCCAACCCCATG

ACCCTGGAAAAGATGGCCTCCCAAGAGC
```

Figure 6.4 Comparison of the coding region of mGK-6, pMK-1 and mGK-1. The nucleotide sequence encoding amino acid residues 150-158 and 206-210 from these three genes are compared. The three nucleotides of mGK-1 that differ from those of mGK-6 and pMK-1 are indicated by stars. The genes mGK-6 and pMK-1 share complete homology over this region.

```
                150                                158       206           210
                SerThrCysLeuAlaSerGlyTrpGly       LysAspThrCysAla
mGK-6      AGCACATGCCTAGCCTCAGGCTGGGGC-------AAAGACACTTGTGCG
pMK-1      AGCACATGCCTAGCCTCAGGCTGGGGC-------AAAGACACTTGTGCG
mGK-1      AGCACATGCCTTGCCTCAGGCTGGGGC-------AAAGACACTTGTAAG
                          *                                **
                                                           Lys
```

analysis, restriction mapping suggests that intron C of mGK-6 is approximately 100 bp long and occurs in the same position as its counterpart from mGK-1. The 5' boundary of intron D separates the sequence coding for amino acids 210 and 211 as demonstrated for mGK-1. Exon 5 from mGK-6 and mGK-2 are both located on 1.1 kb BamHI fragments. Sequences from this intron are approximately 70% homologous with the corresponding regions from mGK-1 and mGK-2. This suggests that the structure of this last exon is conserved between these genes, and thus exon 5 of mGK-6 is probably separated from exon 4 by an intron of 370-380 bp as seen in mGK-1 and mGK-2.

These results demonstrate that the three genes mGK-1, mGK-2 and mGK-6 possess an identical arrangement of introns and exons, although variation is observed in intron size. For example, intron B is approximately 330 bp longer in mGK-6 than in mGK-1.

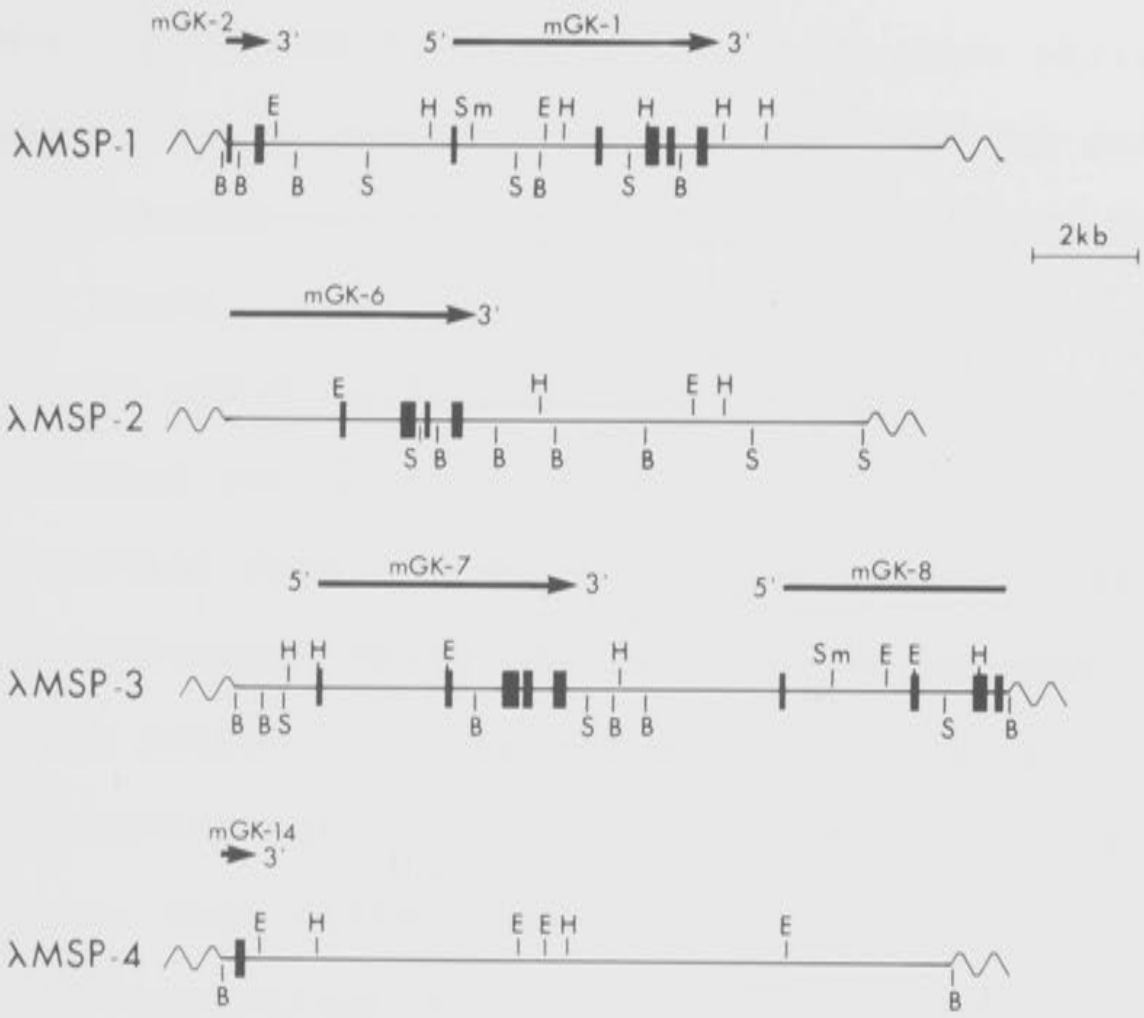
6.2.4 Organisation of the kallikrein gene locus

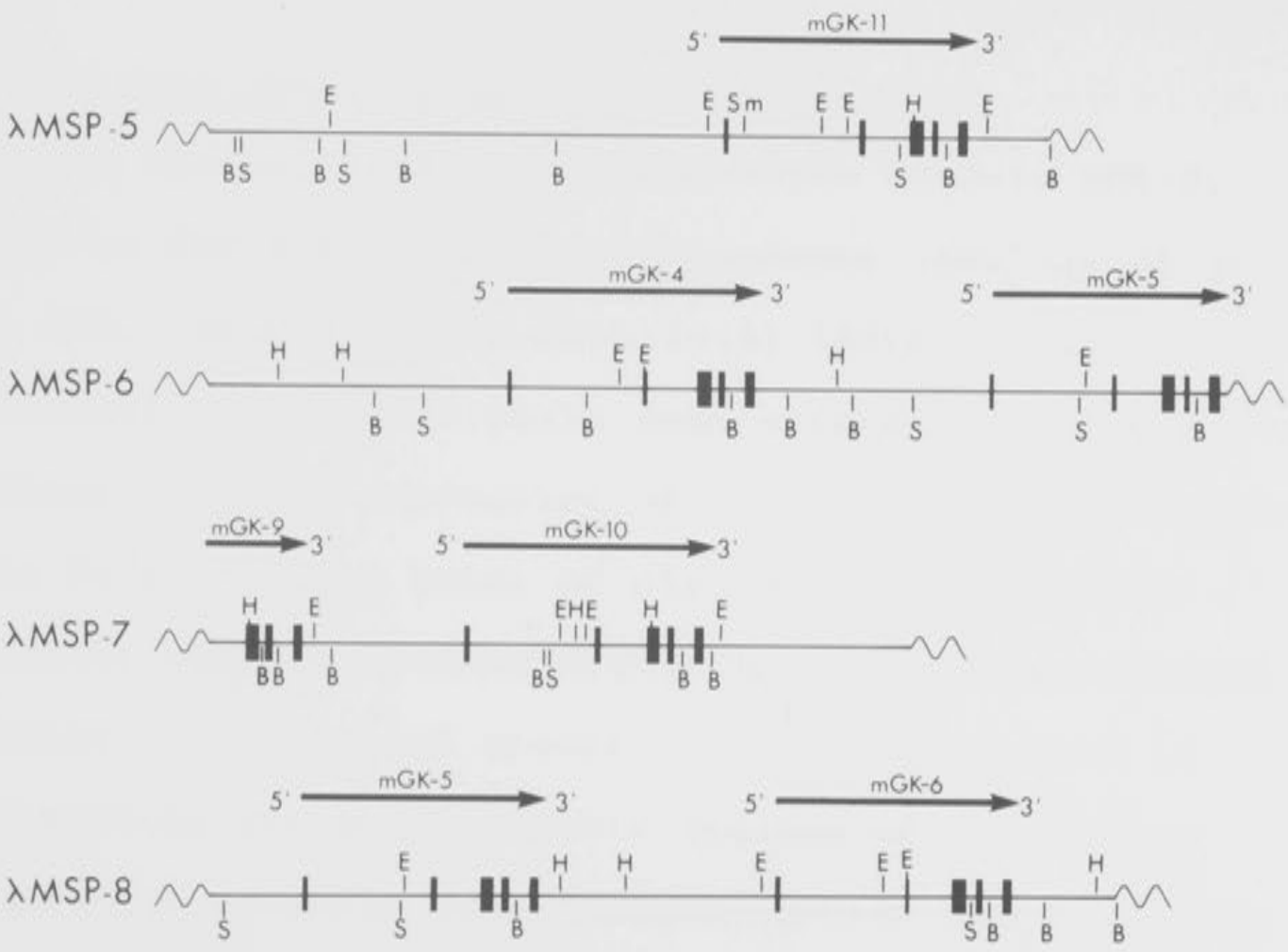
To study the molecular organisation of additional kallikrein genes, a Charon 28 Balb/c mouse embryonic library (supplied by P. Leder) was screened for kallikrein sequences utilising pMK-1 as a hybridisation probe. Approximately 4 X 10^5 phage were screened as described in 2.4.3. Of the 65 clones initially picked, 30 were purified to homogeneity and analysed in detail using the following exon-specific probes.

Exon 1	:	pK-1-HSm.8 - containing exon 1 of mGK-1
Exon 2	:	118 bp TaqI-PvuII fragment from pK-1-B2.6 - containing part of exon 2 from mGK-1
Exon 3, 4, and 5	:	pMK-1
Exon 4	:	pK-2-B.3 - containing exon 4 of mGK-2.
Exon 5	:	pK-1-BH.8 - containing exon 5 of mGK-1

The restriction maps and the location of the kallikrein genes within λMSP-1-2 are illustrated in Figure 6.5 and compared to the other kallikrein gene-containing lambda clones subsequently analysed by B. Evans. The use of exon-specific probes enabled the position of each exon to be determined and the direction of transcription of each gene to be established. It should be noted that the lengths of certain regions of DNA, particularly those of the first intervening sequence, are only approximate. More precise restriction mapping and limited DNA sequence analysis is needed to clarify the exact location of some exons.

Figure 6.5 Restriction maps of cloned DNA fragments containing mouse glandular kallikrein genes. The clones λMSP-1-8 were isolated from mouse genomic libraries using pMK-1 as a hybridisation probe. The characterisation of λMSP-1 and λMSP-2 are described in Chapters 5 and 6, respectively. The positions of exons (denoted by boxes) within the other clones (λMSP-3-8) were mapped by using exon-specific hybridisation probes (see 6.2.4). The direction of transcription of the genes is indicated by the arrows. Nucleotide sequence data indicates that the gene mGK-6 on λMSP-2 is identical to the mGK-6 gene on λMSP-8. Clones λMSP-6 and λMSP-8 overlap. Wavy lines indicate the two arms of the λ Charon 28 vector. Restriction sites: B-BamHI, E-EcoRI, H-HindIII, S-SacI, Sm-SmaI.


mGK-2
3'
5'
mGK-1
3'
λMSP-1
E
H Sm
E H
H
H H
B B B S S B S B
2kb
mGK-6
3'
λMSP-2
E
H
E H
S B B B B S S
5'
mGK-7
3'
5'
mGK-8
λMSP-3
H H
E
H
Sm E E H
B B S B S B B S B
mGK-14
3'
λMSP-4
E H
E E H
E
B B


5'
mGK-11
3'
λMSP-5
E
E Sm
E E
H
E
B S B S B B S B B
5'
mGK-4
3'
5'
mGK-5
3'
λMSP-6
H H
E E
H
E
B S B B B B S S B
mGK-9
3'
5'
mGK-10
3'
λMSP-7
H
E
E H E
H
E
B B B B S B B
5'
mGK-5
3'
5'
mGK-6
3'
λMSP-8
E
H H
E
E E
H
S S B S B B B

Several of these additional clones (λMSP-3, 6, 7 and 8) contain part or all of two kallikrein genes. These linked genes are transcribed off the same strand and are separated by a spacer region of between 3.5 and 5.0 kb.

Sequence analysis of one of the genes present on λMSP-8 (B. Evans, pers. comm.) indicates that it is identical with the mGK-6 kallikrein gene contained on λMSP-2. This assignment was also based on the similarity of the restriction maps of the two clones in this region - the only difference being the presence of an additional EcoRI site in intron A of the mGK-6 gene contained on λMSP-8. Since λMSP-2 was obtained from a Quackenbush mouse genomic library, this difference probably reflects the restriction site polymorphisms such as that described in 4.2.2.

6.2.5 Gene-specific hybridisation probes

All the kallikrein gene sequences (mGK-1, mGK-2, pMK-1) so far analysed possess sequence homology of greater than 80%. Hybridisation data (4.2) indicate that all the mouse kallikrein genes share such extensive sequence homology. If the expression of individual kallikrein genes is to be studied, a means of distinguishing between different members of this highly homologous gene family is essential. The use of specific oligonucleotides representing the more variable regions of the coding sequence as specific hybridisation probes was therefore investigated.

A 19 bp region from the 3' end of the coding region was selected as the most divergent in nucleotide sequence among the three gene sequences available (mGK-1, mGK-2, pMK-1). The sequences from this region of the three genes is compared in Figure 6.6a. Within this 19 bases, both mGK-2 and pMK-1 have a total of 7 base mis-matches with the coresponding sequence from mGK-1. A synthetic oligonucleotide of sequence 5'-CAGGGGTTTTGGGCCAAAG-3', which is complementary to this region of mGK-1, was chemically synthesised by the phosphoramidite method (Beaucage and Caruthers, 1981). Approximately 10 ng of the 19-mer (a gift from J. Coghlan, Howard Florey Institute, Melbourne) was labelled at the 5' terminus to high specific activity with γ-^{32}P-ATP and T4 polynucleotide kinase (2.3.5). This probe was hybridised to a nitrocellulose filter on which EcoRI-cleaved DNA, containing 14 distinct kallikrein genes (see Discussion), had been immobilised (Figure 6.6 b,c). As shown in Figure 6.6c, only the 8.0 kb EcoRI fragment from λMSP-1 hybridised with the labelled oligonucleotide. This restriction fragment contains the appropriate region of mGK-1 which is homologous to the 19-mer. Under the hybridisation conditions used (Legend Figure 6.6) the oligonucleotide is able to specifically detect mGK-1 kallikrein sequences. As previously demonstrated (5.2.5), the mGK-1 gene is expressed in the male submaxillary gland. The 19-mer was shown by northern blot analysis to hybridise to mGK-1 mRNA present in this tissue (data not shown).

Figure 6.6 mGK-1 specific hybridisation probe.

(a) Comparison of the nucleotide sequence of mGK-2, mGK-1 and pMK-1 across the 19 bp region used to design a mGK-1 specific oligonucleotide. This region includes the last two nucleotides of the codon for amino acid residue 256, through to the first two nucleotides of the termination codon. Stars indicate the positions at which the sequence of mGK-2 or pMK-1 differ from that of mGK-1. The sequence of the synthetic 19-mer which is complementary to this region of mGK-1 is shown.

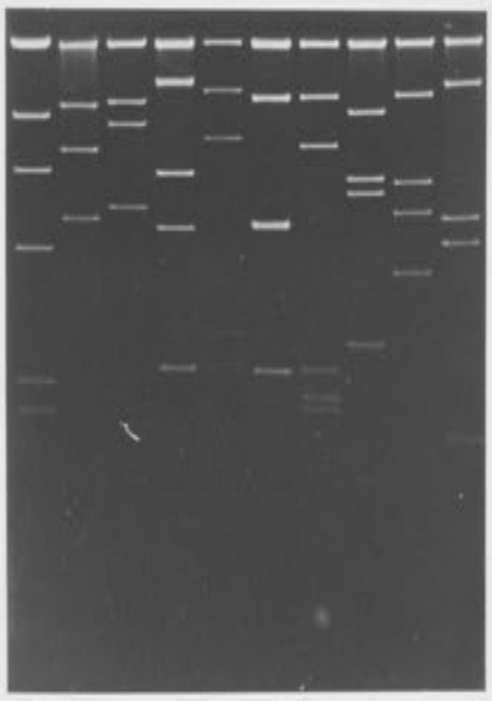
(b) Ethidium bromide stained 1% agarose gel after electrophoresis of EcoRI-cleaved DNA from the λ clones : λMSP-1 (lane 2), λMSP-22 (lane 3), λMSP-8 (lane 4), λMSP-25 (lane 5), λMSP-7 (lane 6), λMSP-5 (lane 7), λMSP-11 (lane 8), λMSP-20 (lane 9), and λMSP-4 (lane 10). These clones contain 14 kallikrein genes (see Figure 6.5 and B. Evans, pers. comm.). The marker DNA is λc1857 DNA cut with HindIII.

(c) Autoradiograph of the Southern blot of the above gel hybridised with the 5'end labelled mGK-1 primer. The nitrocellulose filter was hybridised with the labelled primer for 16 hours in 50% formamide, 3 x SSC at 4°C. The filter was washed twice at 21°C in two changes of 2 x SSC. The only fragment which hybridised was the 8.0 kb EcoRI fragment from λMSP-1.

a

```
mGK-2      5'-CTATGACGAAAAATGCCTA-3'
             *  * **     **    *
mGK-1      5'-CTTTGGCCCAAAACCCCTG-3'
             *  ****     **
pMK-1      5'-CTATGATGAAAAATGCCTG-3'


Primer     5'-CAGGGGTTTTGGGCCAAAG-3'
```

b



c

These experiments demonstrate that by using synthetic oligonucleotides homologous to short stretches of a particular kallikrein gene, the tissue-specific expression of that gene can be studied. Clearly, the isolation of the remaining kallikrein gene sequences is necessary before the uniqueness of such probes can be conclusively established.

6.3 Discussion

The assignment of the mouse kallikrein genes to chromosome seven and their frequent close linkage is consistent with the clustering of these genes into a single locus. Out of the 30 lambda clones analysed to date (this work and B. Evans, pers. comm.) a total of 14 distinct kallikrein genes have been identified. Many of the kallikrein genes are tightly clustered in a head to tail arrangement each separated by an intergenic space of between 3.5 and 5.0 kb. More recently, B. Evans (pers. comm.) has been able to map a series of overlapping lambda clones and demonstrate that at least 4 kallikrein genes (mGK-3, 4, 5 and 6) are tightly linked in the genome. These four genes are encoded on the same strand of DNA and are roughly 4-5 kb apart. However examination of other clusters of kallikrein genes show that some intergenic spaces are considerably larger than observed for this group of four genes.

The strong conservation of restriction sites observed

between different genes and their high degree of cross-hybridisation suggest that sequences in the coding and non-coding regions are highly conserved. The location of EcoRI restriction sites around and within different kallikrein genes is strongly conserved. The different EcoRI fragments from the 14 genes so far analysed, are all represented among the EcoRI fragments detected by genomic blot analysis using pMK-1 as a hybridisation probe (Figure 4.1). Three of these genes contain sequences homologous to pMK-1 on 6.7 kb EcoRI fragments and another three contain hybridising sequences on 2.4 kb EcoRI fragments (B. Evans, pers. comm.). Therefore, the relatively more intense hybridisation of certain EcoRI fragments (e.g: 2.4 and 6.7 kb) observed by genomic blot analysis can be explained by the conservation of EcoRI restriction sites around several distinct genes. By a comparison of the intensity of hybridisation of differently-sized EcoRI restriction fragments (see Figure 4.1) with the prevalence of the same-sized EcoRI fragments among different kallikrein gene isolates, it is possible to estimate that there are between 25 and 30 distinct kallikrein genes in the mouse.

The high degree of sequence homology observed in coding and non-coding regions of different kallikrein genes suggests that these 25-30 genes have recently diverged from a common ancestor. Similar to other multi-gene families, for example α and β globin (Lauer et al., 1980; Efstratiadis et al., 1980); and growth hormone (Seeburg, 1982), the precise boundaries of exon-intron

junctions are conserved between members, although variation is observed in intron size. These data indicate that the introns were present prior to any event of gene amplification and subsequent divergence.

As observed in Chapter 5, functional differences exist between the kallikreins so far analysed. One might postulate that such differences exist between all the kallikreins and that each individual member is involved in the processing of a specific peptide precursor. At present, it is not known whether all the 25-30 kallikrein genes are transcriptionally active. In the submaxillary gland at least 6 members are known to be expressed (γ-NGF, EGF-BP, γ-renin, β-NGF endopeptidase, pMK-1 and mGK-1). The use of synthetic oligonucleotides as gene-specific probes offers a method by which the levels and type of kallikrein gene expression in different tissues can be examined. At present it is known that kallikreins are expressed in the submaxillary gland, kidney, mammary glands, pancreas, stomach, intestine, lung and brain cells (Orstavik et al., 1975; Simson et al., 1978; Orstavik and Glenner, 1978; Orstavik et al., 1979; Peeters et al., 1976; Uchida et al., 1980; Moriwaki et al., 1980 and Werle, 1982). Due to the immunological cross-reactivity of many members of this family it is impossible to ascertain which particular kallikrein protease is being detected. The use of gene-specific probes should resolve this problem.

CHAPTER SEVEN

GENERAL DISCUSSION

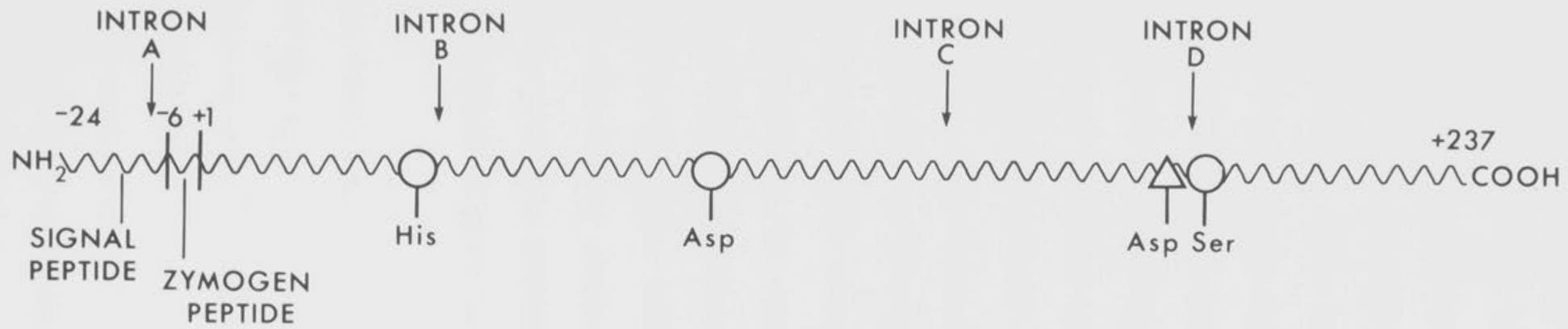
The use of the kallikrein cDNA clone pMK-1 as a hybridisation probe to study the kallikrein family of enzymes at the genomic level has led to the discovery of a very large group of homologous kallikrein genes.

Each of the sequences so far examined, although displaying an overall homology of approximately 75%, differ at amino acids implicated in the substrate specificity of serine proteases. It is possible that such differences may exist between most, if not all, of the kallikrein gene family members. Given the known role of three of these enzymes, namely glandular kallikrein, EGF-BP and γ-NGF, it seems likely that the members of this family fulfill a general role in the specific processing of bioactive peptides. This work supports the previously-suggested hypothesis that kallikrein is only one member of a highly homologous group of enzymes. This suggestion was based on the finding of at least four other kallikrein-like activities in addition to γ-NGF, EGF-BP and glandular kallikrein in the mouse submaxillary gland (Wilson and Shooter, 1979; Bothwell et al., 1979). The work presented in this thesis clearly shows that many of the problems encountered in working with this group of enzymes at the protein level, which have made it difficult even to ascertain how many different kallikrein activities exist, can be obviated by the application of recombinant DNA technology.

Detailed structural analysis of the two kallikrein genes in the genomic clone λMSP-1 demonstrated that these

genes (mGK-1 and mGK-2) are closely linked in the genome. Furthermore, they are greater than 80% homologous in nucleotide sequence in the coding regions and exhibit an identical exon-intron structure. Subsequent studies have shown this to be true for other members of this gene family (B.Evans, pers. comm.). The kallikrein genes each consist of five exons spanning approximately 4.5 kb. The introns separate the coding regions of a preprokallikrein as shown in Figure 7.1. As illustrated, the histidine, serine and aspartic acid residues, which are essential for the formation of a serine protease active site (Young et al., 1978; Krieger et al., 1974), are located on separate exons. Furthermore, the aspartic acid residue (pos.183) which is required for cleavage at basic amino acids (Krieger et al., 1974), a characteristic of these enzymes, is also located on a separate exon. The residues lining the substrate binding pocket which tend to be the more variable regions of the protein are widely distributed throughout the exons (see Figure 5.12). The genomic structure of the related rat trypsin genes is identical to that reported here for kallikrein genes (C. Craik, pers. comm.). These data are consistent with the evolution of a serine protease by the formation of an active site through the joining of different exons - each encoding an amino acid sequence crucial to the final tertiary structure of the catalytic pocket.

Figure 7.1 Position of exon-intron boundaries in the mGK-1 gene. The position of introns A-D are shown in relation to the structure of the preprokallikrein encoded by mGK-1. The three circled residues (His, Asp and Ser) are crucial to the formation of a serine protease catalytic site (Young et al., 1978; Krieger et al., 1974). The aspartic acid residue represented as a triangle confers on a serine protease the ability to cleave at lysine or arginine basic residues (Krieger et al., 1974).


INTRON
A
INTRON
B
INTRON
C
INTRON
D
-24
-6
+1
NH2
+237
COOH
SIGNAL
PEPTIDE
ZYMOGEN
PEPTIDE
His
Asp
Asp
Ser

Elucidation of the complete structure of the mGK-1 gene was necessary to form a firm basis on which other members of this family could be studied. The high degree of cross-hybridization, as observed by Southern blot analysis, suggests that they are roughly 80-85% homologous at the nucleotide level (as are the genes mGK-1, mGK-2 and pMK-1). By a comparison of the information gained from genomic blot analysis with the subsequent data obtained from the analysis of genomic clones (B. Evans, pers. comm.) it is estimated that there are between 25-30 kallikrein genes in the mouse. The fact that all the kallikrein genes are located on chromosome 7 and their frequent close linkage suggests that they may form a tandem array of 25-30 genes. This arrangement is consistent with the organisation of other mammalian multi-gene families e.g. α and β globin (Lauer et al., 1980; Efstratiadis et al., 1980) the light and heavy chain immunoglobulin families (Tonegawa, 1983), and the growth hormone family (Kidd and Saunders, 1982) where highly homologous genes are found in tandem.

In several multi-gene families, e.g. human β-globin genes (Slightom et al., 1981) and the mouse γ-immunoglobulin-constant region genes (Schreier et al., 1981; Ollo and Rougeon, 1983), gene conversion has been suggested as the mechanism by which the set of genes maintains its sequence homogeneity. Gene conversion involves an interaction between gene A and gene B such that

the nucleotide sequence of part or all of gene A becomes identical to that of gene B. By this mechanism, as distinct from unequal crossing-over, gene A and gene B retain their physical locations, although a non-reciprocal alteration of one gene sequence occurs. In the case of the immunoglobulin γ-constant region genes (γ_{2a}, γ_{2b}) it was shown that the two non-allelic genes of Balb/c mice share homologous sequences not present in their allelic counterparts in the C57BL/6 mouse strain (Ollo and Rougeon, 1983). If such gene conversion events were operating in the kallikrein gene locus, one would expect to find extensive nucleotide sequence allelic heterogeneity between different strains of mice. Restriction analysis of the kallikrein gene loci of the mice strains Balb/c, C57BL/6 and Quackenbush indicates very little heterogeneity in their respective kallikrein loci. Obviously, the nucleotide sequences of a variety of kallikrein gene alleles needs to be determined before any gene conversion mechanisms could be implicated. If one assumes, as suggested by the results presented in this thesis, that each of the kallikrein genes has a specific functional role in protein processing, a mechanism of gene conversion would seem detrimental. Gene conversion would only serve to duplicate one gene function whilst destroying another. Indeed, before these questions can be examined, the exact number of <u>functional</u> kallikrein genes needs to be ascertained. One would expect the presence of some

'pseudogenes' as found in most other eukaryotic multi-gene families (Proudfoot, 1980; Sharp, 1983). The use of gene-specific synthetic oligonucleotides, as hybridisation probes on mRNA isolated from a wide range of of tissues, offers the most practical way of determining whether each of the different 25-30 kallikrein genes is expressed.

As described in Chapter 4, the hybridisation obtained with pMK-1 on rat genomic DNA was considerably less intense than that observed for mouse genomic DNA. This result suggests that all the mouse genes are more homologous to each other than to their respective counterparts in the rat. Recently, the complete nucleotide sequence of a cDNA clone encoding rat pancreatic kallikrein was published (Swift et al., 1982). The nucleotide sequence of the rat gene shows 77-80% homology with that of the mouse genes mGK-1, mGK-2 and pMK-1 (which themselves are >80% homologous). The lower extent of hybridisation observed between rat and mouse genes can be explained by the finding that, although these genes share extensive homology, the rat gene lacks the long stretches of homology (up to 50 bp) retained among the mouse genes. Preliminary nucleotide sequence analysis of different rat kallikrein genes has demonstrated that such long stretches of nucleotide sequence homology also exist between distinct rat genes (B. Evans, pers. comm.). Therefore it appears that the mouse genes have maintained more extensive homology with each other than their rat counterparts - this is supported

by the low immunological cross-reactivity observed between rat and mouse kallikrein proteases (Bothwell et al., 1979).

Following the identification of the existence of a multi-gene family, whose members could fulfill a general role in bioactive peptide processing, there are several aspects of the molecular biology of the kallikrein gene family which remain to be investigated in detail. One of the most intriguing problems to be elucidated is the pattern of tissue-specific expression. As outlined in 1.8, kallikrein or kallikrein-like activities have been found in a wide variety of tissues. The isolation and characterisation of all the 25-30 kallikrein genes would aid such an analysis and enable the design of gene-specific oligonucleotides. These could be used as hybridisation probes to detect specific-kallikrein mRNA in different tissues. The construction of cDNA libraries from the mRNA of tissues found to contain kallikrein sequences should facilitate the complete analysis of the different kallikrein genes expressed in such tissues. Once a pattern of tissue-specific expression has been established, the basis for this phenomenon could be investigated at the molecular level.

Since EGF-BP and γ-NGF are members of the same family one can speculate that the other kallikreins may also process a specific bioactive peptide. The most challenging aspect of this work is to determine the specific roles played by each kallikrein gene in the activation and

processing of a variety of physiologically important polypeptides. The results gained from the tissue studies should give clues as to which polypeptide each member is processing, since it is probable that the processed polypeptide would be co-ordinately expressed with its respective processing enzyme. The use of eukaryotic gene expression systems as a means of obtaining highly purified kallikreins should be investigated. If successful, the purified proteases could be tested in vitro for the ability to process a wide variety of prohormones and growth factors.

In addition, experiments could be conducted to investigate the molecular basis of co-ordinate gene expression. Since in vivo, the β-NGF and γ-NGF proteins are found in stoichiometric amounts (Server and Shooter, 1976), their expression is presumably co-ordinately regulated. Gene transfer experiments involving the introduction of the genes coding for β-NGF and γ-NGF into eukaryotic cells coupled with in vitro mutagenesis of 'control' regions may provide insights into the molecular basis for co-ordinate gene expression.

REFERENCES

Adams, J.M. (1977) Messenger RNA. In "The Ribonucleic Acids", eds. Stewart P.R., and D.S. Letham, Springer-Verlag, New York, Heidelberg, Berlin, pp 81-128.

Aloe, L. and R. Levi-Montalcini (1979) Nerve growth factor induced transformation of immature chromaffin cells *in vivo* into sympathetic neurons: effects of antiserum to nerve growth factor. Proc. Natl. Acad. Sci. USA 76, 1246-1250.

Angeletti, R.H. and R.A. Bradshaw (1971) Nerve growth factor from mouse submaxillary gland: Amino acid sequence. Proc. Natl. Acad. Sci. USA 68, 2417-2420.

Angeletti, R.H., R.A. Bradshaw and G.L. Marshall (1974) The synthesis and characterisation of the amino-terminal octapeptide of mouse nerve growth factor. Int. J. Protein Res. 6, 321-328.

Aviv, H. and P. Leder (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69, 1408-1412.

Attardi, D.G., R. Levi-Montalcini, B.S. Wenger and P.U. Angeletti (1969) Submaxillary gland of mouse: Effects of a fraction on tissues of mesodermal origin *in vitro*. Science 150, 1307-1309.

Beaucage, S.L. and M.H. Caruthers (1981) Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. Tetra. Lett. 22, 1859-1862.

Benoist, C., K. O'Hare, R. Breathnach and P. Chambon (1980) The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Benton, W.D. and R.W. Davis (1979) Screening λgt recombinant clones by hybridisation to single plaques *in situ*. Science 196, 180-182.

Berger, E.A. and E.M. Shooter (1977) Evidence for pro-β-nerve growth factor, a biosynthetic precursor to β-nerve growth factor. Proc. Natl. Acad. Sci. USA 74, 3647-3651.

Birnbaum, J.E., T.M. Sapp and J.B. Moore Jr. (1976) Effects of reserpine, epidermal growth factor, and cyclic nucleotide modulation on epidermal mitosis. J. Invest. Dermatol. 66, 313-318.

Bode, W., Z. Chen, K. Bartles, C. Kutzbach, G. Schmidt-Kastner and H. Bartunik (1983) Refined 2A X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. J. Mol. Biol. 164, 237-282.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heynecker, H.W. Boyer, J.H. Crosa and S. Falkow (1977) Construction and characterisation of new cloning vehicles, II. A multipurpose cloning system. Gene 2, 95-113.

Bothwell, M.A., W.H. Wilson and E.M. Shooter (1979) The relationship between glandular kallikrein and growth factor-processing proteases of mouse submaxillary gland. J. Biol. Chem. 254, 7287-7294.

Breathnach, R. and P. Chambon (1981) Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Brown, D.D. (1981) Gene expression in eukaryotes. Science 211, 667-674.

Burton, L.E., W.H. Wilson and E.M. Shooter (1978) Nerve growth factor in mouse saliva. J. Biol. Chem. 253, 7807-7812.

Burton, L.E. and E.M. Shooter (1981) The molecular basis of the heterogenity of the γ-subunit of 7S nerve growth factor. J. Biol. Chem. 256, 11,011-11,017.

Carpenter, G. and S. Cohen (1979) Epidermal growth factor. Ann. Rev. Biochem. 48, 193-216.

Catterton, W.Z., M.B. Escobedo, W.R. Sexson, M.E. Gray, H.W. Sundell and M.T. Stahlman (1979) Effect of epidermal growth factor on lung maturation in fetal rabbits. Pediatr. Res. 13, 104-108.

Chaconas, G. and J.H. van de Sande (1980) 5'-^{32}P labelling of RNA and DNA restriction fragments. Meth. Enzymol. 65, 75-85.

Chrétien, M. and N.G. Seidah (1981) Chemistry and biosynthesis of pro-opiomelanocortin. Molec. Cell Bioch. 34, 101-127.

Clewell, D.B. (1972) Nature of ColE1 plasmid replication in Escherichia coli in the presence of chloramphenicol. J. Bacteriol. 110, 667-676.

Cooke, N.E., D. Coit, J. Shine, J.D. Baxter and J.A. Martial (1981) Human prolactin: cDNA structural analysis and evolutionary comparisons. J. Biol. Chem. 256, 4007-4016.

Corden, J., B. Wasylyk, P. Buchwalder, P. Sassone-Corsi, C. Kedinger and P. Chambon (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414.

Cox D.R., J.A. Sawicki, D. Yee, E. Appella and C.J. Epstein (1982) Assignment of the gene for β_2-microglobulin (B2m) to mouse chromosome 2. Proc. Natl. Acad. Sci. USA 79, 1930-1934.

Darnell, J.E. (1982) Variety in the level of gene control in eukaryotic cells. Nature 297, 365-371.

Davie, E.W. and K. Fujikawa (1975) Basic mechanisms in blood coagulation. Ann. Rev. Biochem. 44 799-829.

Dayhoff, M.O. (ed) (1978) Atlas of protein sequence and structure, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Silver Spring, MD.

Docherty, K. and D.F. Steiner (1982) Post-translational proteolysis in polypeptide hormone biosynthesis. Ann. Rev. Physiol. 44, 625-638.

Docherty, K., R.J. Carroll and D.F. Steiner (1982) Conversion of proinsulin to insulin: Involvement of a 31,500 molecular weight thiol protease. Proc. Natl. Acad. Sci. USA 79, 4613-4617.

Efstratiadis, A., J.W. Pasakony, T. Maniatis, R.M. Lawn, C. O'Connell, R.A. Spritz, J.K. DeRiel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.E. Blechl, O. Smithies, F.E. Baralle, C.C. Shoulders and N.J. Proudfoot (1980) The structure and evolution of the human β-globin gene family. Cell 21, 653-668.

Ekfors, T.O. and V.K. Hopsu-Havu (1971) Immunofluorescent localisation of trypsin-like esteropeptidases in the mouse submandibular gland. J. Histochem. 3, 415-420.

Fiddes, J. and H.M. Goodman (1979) Isolation, cloning and sequence analysis of the cDNA for the α-subunit of human chorionic gonadotropin. Nature 281, 351-356.

Fiddes, J. and H.M. Goodman (1980) The cDNA for the β-subunit of human chorionic gonadotropin suggest evolution of a gene by read through into the 3'-untranslated region. Nature 286, 684-687.

Fiedler, F. and H. Fritz (1981) Striking structural similarity of pig pancreatic kallikrein, mouse nerve growth factor γ-subunit and rat tonin. Hoppe-Seyler's Z. Physiol. Chem. 362, 1171-1175.

Fletcher, D.J., B.D. Noe, G.E. Bauer and J.P. Quigley (1980) Characterisation of the conversion of a somatostatin precursor to somatostatin by islet secretory granules. Diabetes 29, 593-599.

Fletcher, D.J., J.P. Quigley, G.E. Bauer and B.D. Noe (1981) Characterisation of proinsulin- and proglucagon-converting activities in isolated islet secretory granules. J. Cell Biol. 90, 312-322.

Frey, P., R. Forand, T. Maciag and E.M. Shooter (1979) The biosynthetic precursor of epidermal growth factor and the mechanism of its processing. Proc. Natl. Acad. Sci. USA 76, 6294-6298.

Gabbay, K.H., R.M. Bergenstal, J. Wolff, M.E. Mako and A.H. Rubenstein (1979) Famial hyperproinsulinemia: Partial characterisation of circulating proinsulin-like material. Proc. Natl. Acad. Sci. USA 76, 2881-2885.

Greene, L.A., E.M. Shooter and S. Varon (1969) Subunit interaction and enzymatic activity of mouse 7S nerve growth factor. Biochemistry 8, 3735-3741.

Greene, L.A., S. Varon, A. Piltch and E.M. Shooter (1971) Substructure of the β-subunit of mouse 7S nerve growth factor. Neurobiology 1, 37-48.

Gresik, E. and T. Barka (1977) Immunocytochemical localisation of epidermal growth factor in mouse submandibular gland. J. Histochem. Cytochem. 25, 1027-1035.

Gross-Bellard, M., P. Oudet and P. Chamon (1973) Isolation of high-molecular-weight DNA from mammalian cells. Eur. J. Biochem. 36, 32-38.

Grosschedl, R. and M.L. Birnstiel (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA 77, 1432-1436.

Grosveld, G.C., E. de Boer, C.K. Shewmaker and R.A. Flavell (1982) DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature 295, 120-126.

Grunstein, M. and J. Wallis (1979) Colony hybridisation. Meth. Enzymol. 68, 379-389.

Guerry, P., D.J. Le Blanc and S. Falkow (1973) General method for the isolation of plasmid deoxyribonucleic acid. J. Bacteriol 116, 1064-1066.

Hagenbuchle, O., R. Bovery and R.A. Young (1980) Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell 21, 179-187.

Harris-Warrick, R.M., M.A. Bothwell and E.M. Shooter (1980) Subunit interactions inhibit the binding of β nerve growth factor to receptors on embryonic chick sensory neurons. J. Biol. Chem. 255, 11,284-11,289.

Hartley, B.S. and D.M. Shotton (1971) Pancreatic elastase. In "The enzymes", vol. 3. Ed. Boyer, P.D., Academic Press, New York and London, pp. 323-353.

Heidecker, G., J. Messing and B. Gronenborn (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene 10, 69-73.

Hermodson, M.A., L.H. Ericsson, H. Neurath and K.A. Walsh (1973) Determination of the amino acid sequence of porcine trypsin by sequenator analysis. Biochemistry 12, 3146-3153.

Hiramatsu, M., K. Hatakeyama, M. Kumegawa, T. Yajima and N. Minami (1981) Genetic variation in esteroproteases in the mouse submandibular gland. Experientia 37, 1068-1069.

Ho, P.C., W.H. Davis, J.H. Elliot and S. Cohen (1974) Kinetics of corneal epithelial regeneration and epidermal growth factor. Invest. Ophthalmol. 13, 804-809.

Hobart, P., R. Crawford, L.P. Shen, R. Pictet and W.J. Rutter (1980) Cloning and sequence analysis of cDNAs encoding two distinct somatostatin precursors found in the endocrine pancreas of the anglerfish. Nature 288, 137-141.

Hohn, B. (1979) In vitro packaging of λ and cosmid DNA. Meth. Enzymol. 68, 299-309.

Hudson, P., J. Haley, M. Cronk, J. Shine and H. Niall (1981a) Molecular cloning and characterisation of cDNA sequences coding for rat relaxin. Nature 291, 127-131.

Hudson, P., J. Penschow, J. Shine, G. Ryan, H. Niall and J. Coghlan (1981b) Hybridisation histochemistry: use of recombinant DNA as a "homing probe" for tissue localisation of specific mRNA populatiions. Endocrinology 108, 353-356.

Hugli, T.E. (1975) Serum anaphylatoxins: formation, characterisation and control. In "Proteases and Biological Control", vol 2. Eds. Reich, E., D.B. Rifkin and Shaw, E., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 273-290.

Kakidani, H., Y. Furutani, H. Takahashi, M. Noda, Y. Morimoto, T. Hirose, M. Asai, S. Inayama, S. Nakanishi and S. Numa (1982) Cloning and sequence analysis of cDNA for porcine β-neo-endorphin/dynorphin precursor. Nature 298, 245-249.

Kaplan, B.B., S.L. Bernstein, A.E. Gioio (1979) An improved method for the rapid isolation of brain ribonucleic acid. J. Biochem. 183, 181-184.

Kidd, V.J. and G.F. Saunders (1982) Linkage arrangement of human placental lactogen and growth hormone genes. J. Biol. Chem. 157, 10,673-10,680.

Krieger, M., L.M. Kay and R.M. Stroud (1974) Structure and specific binding of trypsin: comparison of inhibited derivatives and a model for substrate binding. J. Mol. Biol. 83, 209-230.

Land, H., G. Schutz, H. Schmale and D. Richter (1982) Nucleotide sequence of cloned cDNA encoding bovine arginine vasopressin-neurophysin II precursor. Nature, 295, 299-303.

Land, H., M. Grez, S. Ruppert, H. Schmale, M. Rehbein, D. Richter and G. Schutz (1983) Deduced amino acid sequence from the bovine oxytocin-neurophysin I precursor cDNA. Nature 302, 342-344.

Lauer, J., C-K.J. Shen and T. Maniatis (1980) The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. Cell 20, 119-130.

Lazure, C., N.G. Seidah, G. Thibault, J. Genest and M. Chrétien (1981) Amino acid sequence comparison of rat tonin with nerve growth factor γ-subunit and with serine proteases. In Peptides: Synthesis, structure and function. Proceedings of the VIIth American Peptide Symposium. Eds. Rich, D.H. and M.R. Gross, Pierce Chemical, Rockford, 1L, pp. 517-520.

Legerski, R.J., J.L. Hodnett and H.B. Gray (1978) Extracellular nucleases of Pseudomonas Bal31, III. Use of the double-stranded deoxyriboexonuclease activity as the basis of a convenient method for mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. 5, 1445-1463.

Lerner, M.R., J.A. Boyle, S.M. Mount, S.L. Wolin and J.A. Steitz (1980) Are snRNPs involved in splicing? Nature 283, 220-224.

Levi-Montalcini, R. and P.U. Angletti (1968) Nerve growth factor. Physiol. Rev. 48, 534-569.

Levi-Montalcini, R. (1976) The nerve growth factor: its role in growth differentiation and function of the sympathetic adrenergic neuron. Prog. Brain Res. 45, 235-258.

Liu, C-P., P.W. Tucker, J.F. Mushinski and F.R. Blattner (1980) Mapping of heavy chain genes for mouse immunoglobulins M and D. Science 209, 1348-1353.

Loh, Y.P. and H. Gainer (1982) Characterisation of pro-opiocortin-converting activity in purified secretory granules from rat pituitary neurointermediate lobe. Proc. Natl. Acad. Sci. USA 79, 108-112.

MacDonald, R.J., S.J. Stary and G.H. Swift (1982) Two similar nonallelic rat pancreatic trypsinogens. J. Biol. Chem. 257, 9724-9732.

Maniatis, T., R.C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G.K. Sim and A. Efstratiadis (1978) The isolation of structural genes from libraries of eukaryotic DNA. Cell 15, 687-701.

Maroux, S., J. Baratti and P. Desuvelle (1971) Purification and specificity of porcine enterokinase. J. Biol. Chem. 246, 5031-5039

Maxam, A.M. and W. Gilbert (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. Enzym. 65, 499-560.

Messing, J., B. Gronenborn, B. Müller-Hill and P.H. Hofschneider (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a Hind III fragment of the lac regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. USA 74, 3642-3646.

Messing, J., R. Crea and P.H. Seeburg (1981) A system for shotgun DNA sequencing. Nucleic Acids Res. 9, 309-321.

Messing, J. and J. Vieira (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269-276.

Mobley, W.C., A. Schenker and E.M. Shooter (1976) Characterisation and isolation of proteolytically modified nerve growth factor. Biochemistry 15, 5543-5552.

Moriwaki, C., H. Fujimora, Y. Toyona and T. Nagai (1980) Studies on kallikreins. V. Purification and characterisation of rat intestinal kallikrein. Chem. Pharm. Bull., (Tokyo) 28, 3612-3620.

Morrison, D.A. (1979) Transformation and preservation of competent bacterial cells by freezing. Meth. Enzymol. 68, 326-331.

Murray, N.E., W.J. Brammar and K. Murray (1977) Lambdoid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. 150, 53-61.

Nakanishi, S., A. Inowe, T. Kita, M. Nakamura, A.C.Y. Chang, S.N. Cohen and S. Numa (1979) Nucleotide sequence of cloned cDNA for bovine corticotropin-β-lipotropin precursor. Nature 278, 423-427.

Naughton, M.A., J. Koch, H. Hoffman, V. Bender, H. Hagopian and E. Hamilton (1969) Isolation and activity of a thymocyte-transforming factor from the mouse submaxillary gland. Exp. Cell. Res. 57, 95-103.

Neurath, H. and G.H. Dixon (1957) Structure and activation of trypsinogen and chymotrypsinogen. Fed. Proc. 16, 791-801.

Neurath, H. and K.A. Walsh (1976) Role of proteolytic enzymes in biological regulation (a review). Proc. Natl. Acad. Sci. USA 73, 3825-3832.

Noda, M., Y. Teranishi, H. Takahashi, M. Toyosato, M. Notake, S. Nakanishi and S. Numa (1982) Isolation and structural organisation of the human preproenkephalin gene. Nature 297, 431-434.

Ollo, R. and F. Rougeon (1983) The role of gene conversion in the generation of polymorphism in multigenic families: generation of mouse immunoglobulin γ_{2a} chain alleles by differential gene conversion by γ_{2b} chain gene. Cell 32, 515-523.

Orstavik, T.B., R. Brandtzaeg, K. Nustad and K.M. Halvorsen (1975) Cellular localisation of kallikrein in the rat submandibular and sublingual salivary glands: immunofluorescence tracing related to histological characteristics. Acta. Histochem. 54, 183-192.

Orstavik, T.B. and G.G. Glenner (1978) Localisation of kallikrein and its relation to other trypsin-like esterases in the rat pancreas. A comparison with the submandibular gland. Acta. Physiol. Scand. 103, 384-393.

Orstavik, T.B., K. Nustad and P. Brandtzaeg (1979) Origin of kallikrein in rat and human exocrine glands and kidney. Clin. Sci. 57, 239-241.

Orstavik, T.B. (1980) The kallikrein-kinin system in exocrine organs. J. Histochem. Cytochem. 28, 881-889.

Pasquini, F., A. Petris, G. Sbaragua, R. Scopelliti, G. Cenci and L. Frati (1974) Biological activities in the granules isolated from the mouse submaxillary gland. Exp. Cell Res. 86, 233-236.

Pattison, S.E. and M.F. Dunn (1975) On the relationship of zinc ion to the structure and function of the 7S nerve growth factor protein. Biochemistry 14, 2733-2739.

Peacock, A.C. and C.W. Dingham (1967) Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. Biochemistry 6, 1818-1827.

Peeters, G., R. Verbeke, A. Houvenaghel and R. Reynaert (1976) Isolation of kallikrein from mammary gland of cows. Q.J. Exp. Physiol. 61, 1-14.

Poe, M., J.K. Wu, J.R. Florance, J.A. Rodkey, C.D. Bennett and K. Hoogsteen (1983) Purificatiion and properties of renin and γ-renin from the mouse submaxillary gland. J. Biol. Chem. 258, 2209-2216.

Porcelli, G., C. Cozzari, M. DiIorio, R.H. Croxatto and P.N. Angeletti (1976) Isolation and partial characterization of a kallikrein from mouse submaxillary glands. J. Biochem. 25, 337-348.

Proudfoot, N.J. and C.G. Brownlee (1976) 3'non coding region sequences in eukaryotic messenger RNA. Nature 263, 211-214.

Proudfoot, N.J. (1980) Pseudogenes. Nature 286, 840-841.

Quinto, C., M. Quiroga, W.F. Swain, W.C. Nikovits, D.N. Standring, R.L. Pictet, P. Valenzuela and W.J. Rutter (1982) Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterisation of the gene. Proc. Natl. Acad. Sci. USA 79, 31-35.

Richards, R.I., D.F. Catanzaro, A.J. Mason, B.J. Morris, J.D. Baxter and J. Shine (1982) Mouse glandular kallikrein genes. J. Biol. Chem. 257, 2758-2761.

Robbins, D.C., P.M. Blix, A.H. Rubenstein, Y. Kanazawa, K. Kosaha and H.S. Tager (1981) A human proinsulin variant at arginine 65. Nature 291, 679-681.

Roberts, J.L., P.H. Seeburg, J. Shine, E. Herbert, J.D. Baxter and H.M. Goodman (1979) Corticotropin and β-endophin: Construction and analysis of recombinant DNA complementary to mRNA for the common precursor. Proc. Natl. Acad. Sci. USA 76, 2153-2157.

Sanger, F., S. Nicklen and A.R. Coulson (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Savage, C.R. and S. Cohen (1972) Epidermal growth factor and a new derivative: rapid isolation procedures and biological and chemical characterisation. J. Biol. Chem. 247, 7609-7611.

Schachter, M., B. Maranda and C. Moriwaki (1978) Localisation of kallikrein in the coagulating and submandibular glands of the guinea pig. J. Histochem. Cytochem. 26, 318-321.

Schachter, M. (1980) Kallikreins (Kininogenases) - A group of serine proteases with bioregulatory actions. Pharmacol. Rev. 31, 1-17.

Schenkein, I., M. Levy, E.C. Franklin and B. Frangione (1977) Proteolytic enzymes from the mouse submaxillary gland. Specificity restricted to arginine residues. Arch. Biochem. Biophys. 182, 64-70.

Schreier, P.H. and R. Cortese (1979) A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. 129, 169-172.

Schreier, P.H., A.L.M. Bothwell, B. Mueller-Hill and D. Baltimore (1981) Multiple differences between the nucleic acid sequences of the $IgG2a^a$ and $IgG2a^b$ alleles of the mouse. Proc. Natl. Acad. Sci. USA 78, 4495-4499.

Schwab, M.E., K. Stockel and H. Thoenen (1976) Immunocytochemical localisation of nerve growth factor (NGF) in the submandibular gland of adult mice by light and electron microscopy. Cell Tissue Res. 169, 289-299.

Scott, J., M. Selby, M. Urdea, M. Quiroga, G.I. Bell and W.J. Rutter (1983) Isolation and nucleotide sequence of a cDNA encoding the precursor of mouse nerve growth factor. Nature 302, 538-540.

Seeburg, P.H., J. Shine, J.A. Martial, J.D. Baxter and H.M. Goodman (1977) Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. Nature 270, 486-494.

Seeburg, P.H. (1982) The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone. DNA 1, 239-249.

Server, A.C. and E.M. Shooter (1976) Comparison of the arginine esteropeptidases associated with the nerve and epidermal growth factors. J. Biol. Chem. 251, 165-173.

Sharp, P.A., A.J. Berk and S.M. Berget (1980) Transcription maps of adenovirus. Meth. Enzymol. 65, 750-768.

Sharp, P.A. (1983) Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301, 471-472.

Shine, J., P.H. Seeburg, J.A. Martial, J.D. Baxter and H.M. Goodman (1977) Construction and analysis of recombinant DNA for human chorionic somatomammotropin. Nature 270, 494-499.

Shooter, E.M., P. Frey, P.W. Gunning, G.E. Landreth, A. Sutter and B.A. Yanker (1982) Modulation of neuronal function by nerve growth factor. In "Molecular Genetic Neuroscience" Eds. Schmitt, F.O., S.J. Bird and F.E. Bloom, Raven Press, New York, pp. 425-435.

Silverman, R.E. (1977) Ph.D. thesis, Washington University, St. Louis.

Simson, J.A.V., J. Chao and H.S. Margolius (1978) Cytochemical localisation and secretagogue-induced release of kallikrein and nerve growth factor from rodent salivary glands. Anat. Rec.. 190 542-543.

Skow, L.C. (1978) Genetic variation at a locus (TAM-1) for submaxillary gland protease in the mouse and its location on chromosome 7. Genetics 90, 713-724.

Sleigh, M.J., G.W. Both and G.G. Brownlee (1979) A new method for the size estimation of the RNA genome segments of influenza virus. Nucleic Acids Res. 6, 1309-1321.

Slightom, J.L., A.E. Blechl and O. Smithies (1980) Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21, 627-638.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Molec. Biol. 98, 503-517.

Stach, R.W. and E.M. Shooter (1980) Cross-linked 7S nerve growth factor is biologically inactive. J. Neurochem. 34, 1499-1505.

Stach, R.W., P.F. Pignatti, M.E. Baker and E.M. Shooter (1980) The characterisation of the α-subunits of 7S nerve growth factor. J. Neurochem. 34, 850-855.

Steiner, D.F. and P.E. Oyer (1967) The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. Proc. Natl. Acad. Sci. USA 57, 473-480.

Steiner, D.F., W. Kemmler, H.S. Tager and J.D. Peterson (1974) Proteolytic processing in the biosynthesis of insulin and other proteins. Fed. Proc. 33, 2105-2115.

Steiner, D.F. (1982) Proteolytic processing of secretory proteins. In "Molecular Genetic Neuroscience", Eds. Schmitt, F.O., S.J. Bird and F.E. Bloom, Raven Press, New York, pp. 149-159.

Stroud, R.M., M. Krieger, R.E. Koepell, A.A. Kossiakoff and J.L. Chambers (1975) Structure-function relationships in the serine proteases. In "Proteases and Biological Control", vol 2. Eds. Reich, E., D.B. Rifkin and E. Shaw, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 13-32.

Sun, A.M., B.J. Lin and R.E. Haist (1973) Studies on the conversion of proinsulin to insulin in the isolated islets of Langerhans in the rat. Can. J. Physiol. Pharmacol. 51, 175-182.

Swift, G.H., J-C., Dagorn, P.L. Ashley, S.W. Cummings and R.J. MacDonald (1982) Rat pancreatic kallikrein mRNA: Nucleotide sequence and amino acid sequence of the encoded preproenzyme. Proc. Natl. Acad. Sci. USA 79, 7263-7267.

Taylor, J.M., S. Cohen and W.M. Mitchell (1970) Epidermal growth factor- high and low molecular weight forms. Proc. Natl. Acad. Sci. USA 67, 164-171.

Taylor, J.M., W.M. Mitchell and S. Cohen (1974) Characterisation of the binding protein for epidermal growth factor. J. Biol. Chem. 249, 2188-2194.

Taylor, J.M., R. Illmensee and J. Summers (1976) Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. Biochim. Biophys. Acta. 442, 324-330.

Thoenen, H. and Y.-A. Barde (1980) Physiology of nerve growth factor. Physiol. Rev. 60, 1284-1335.

Thomas, P.S. (1980) Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose paper. Proc. Natl. Acad. Sci. USA 77, 5201-5205.

Thomas, K.A., N.C. Baglan and R.A. Bradshaw (1981) The amino acid sequence of the γ-subunit of mouse submaxillary gland 7S nerve growth factor. J. Biol. Chem. 256, 9156-9166.

Tonegawa, S. (1983) Somatic generation of antibody diversity. Nature 302, 575-581.

Tschesche, H., G. Mair, G. Godfec, F. Fiedler, W. Ehret, C. Hirschauer, M. Lemon, H. Fritz, G. Schmidt-Kastner and C. Kutzbach (1979) The primary structure of porcine glandular kallikreins. In: Kinins II: Biochemistry, pathophysiology and clinical aspects. Eds. Fujii, S., H. Moriya and T. Suzuki, Plenum Press, New York, pp. 245-260.

Tu, C.-P.D. and S.N. Cohen (1980) 3'end labelling of DNA with α-^{32}P-cordycepin-5'-triphosphate. Gene 10, 177-182.

Uchida, K., M. Nininobe, H. Kato and S. Fujii (1980) Purification and properties of rat stomach kallikrein. Biochim. Biophys. Acta. 614, 501-510.

Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W.J. Rutter and H.M. Goodman (1977) Rat insulin genes: construction of plasmids containing the coding sequences. Science 196, 1313-1319.

Uy, R. and F. Wold (1977) Post-translational covalent modification of proteins. Science 198, 890-896.

Walter, P. and G. Blobel (1980) Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA 77, 7112-7116.

Walter, P. and G. Blobel (1981) Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in vitro-assembled polysomes synthesising secretory protein. J. Cell Biol. 91, 551-556.

Weimar, V.L. and K.H. Haraguchi (1975) A potent new mesodermal growth factor from mouse submaxillary gland. A quantitative, comparative study with previously described submaxillary gland growth factors. Physiol Chem. Phys. 7, 7-21.

Werle, E. (1972) History of kallikrein and some aspects of its chemistry and physiology. In "Kininogenases-Kallikrein", Vol. 1. Eds. Haberland, G.L. and J.W. Rohen, Stuttgart, New York, F.K. Schattauer Verlag, pp. 7-22.

Wilson, W.H. and E.M. Shooter (1979) Structural modification of the NH_2 terminus of nerve growth factor. J. Biol. Chem. 254, 6002-6009.

Womack, J.E. (1980) Biochemical loci of the mouse (mus musculus). In "Genetic Maps". Ed. O'Brien, S.J. N.I.H., Bethesda, Maryland, pp. 218-224.

Young, C.L., W.C. Barker, C.M. Tomaselli and M.O. Dayhoff (1978) In "Atlas of protein sequence and structure", Vol. 5. Ed. Dayhoff, M.O., National Biomedical Research Foundation, Silver Spring, MD, pp. 73-93.